

INT-1740

v. 2

CEPAL/CELADE (1740)

v. 2

NACIONES UNIDAS  
CENTRO LATINOAMERICANO DE DEMOGRAFÍA (CELADE)

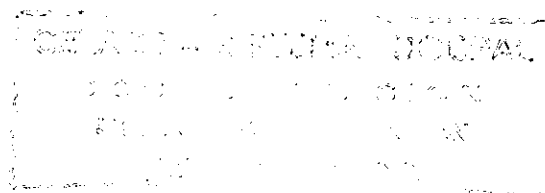


**XX CURSO REGIONAL INTENSIVO  
DE ANÁLISIS DEMOGRÁFICO  
1997**

**1. MÉTODOS CUANTITATIVOS  
ESTADÍSTICAS: MATERIAL DOCENTE**

**MATERIAL DOCENTE**

*(Para uso exclusivo de los alumnos)*



Santiago de Chile

• •

• •

•

•

CURSO INTERNACIONAL INTENSIVO DE  
DEMOGRAFIA



ESTADISTICA

Organiza: Facultad de Ciencias Básicas  
Universidad de Antofagasta.

Patrocinan:

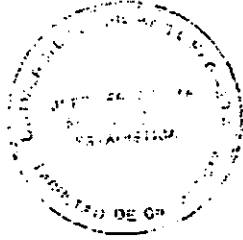


UNIVERSIDAD DE ANTOFAGASTA  
FACULTAD DE CIENCIAS BÁSICAS

INSTITUTO NACIONAL DE ESTADÍSTICA



CENTRO LATINOAMERICANO DE DEMOGRAFÍA



CURSO INTERNACIONAL INTENSIVO DE  
D E M O G R A F I A

APUNTES DE

ESTADÍSTICA

Mg. Jimmy Reyes Rocabado  
Departamento de Matemáticas  
Facultad de Ciencias Básicas  
Universidad de Antofagasta

Mg. Juan Duarte Vargas  
Departamento de Matemáticas  
Facultad de Ciencias Básicas  
Universidad de Antofagasta

UNIVERSIDAD DE ANTOFAGASTA  
FACULTAD DE CIENCIAS BASICAS

INSTITUTO NACIONAL DE ESTADISTICA



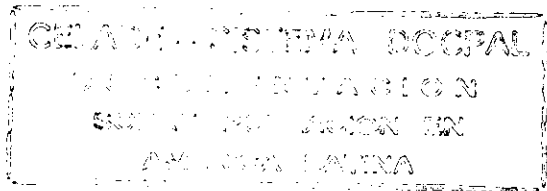
CENTRO LATINOAMERICANO DE DEMOGRAFIA



CURSO INTERNACIONAL INTENSIVO DE  
D E M O G R A F I A

APUNTES DE

ESTADISTICA



Mg. Jimmy Reyes Rocabado  
Departamento de Matemáticas

Mg. Juan Duarte Vargas  
Departamento de Matemáticas

#### IV TASAS, RAZONES Y PROPORCIONES

##### Cifras absolutas y frecuencias relativas

Las estadísticas que resultan de las tabulaciones de diferentes tipos de datos (nacimientos, defunciones, casos de enfermedad, consultas, egresos hospitalarios, etc.) proporcionan números absolutos que son muchas veces utilizables directamente en Salud Pública. Por ejemplo, el número de consultas otorgadas en un consultorio externo permite al administrador en salud estimar la cantidad de recursos necesarios para dar una atención suficiente; el número de nacimientos es un dato valioso para los programas de atención materno-infantil; el número de egresos de un hospital muestra el volumen de hospitalización y sirve para calcular costos y rendimientos.

Sin embargo, a pesar de la importancia de las cifras absolutas, son las frecuencias relativas las que tienen una mayor utilidad. Bajo esta denominación se incluyen las tasas, proporciones, porcentajes y simples razones. Las frecuencias relativas tienen la ventaja de facilitar la presentación de las relaciones que existen entre dos o más datos y hacer más sencilla la comparación de resultados.

##### 1. Razones

Son cocientes entre dos cantidades de igual o distinta naturaleza. Indican cuantas veces sucede el hecho que está en el numerador con respecto al hecho que está en el denominador.

Ejemplo : Razón de masculinidad =  $\frac{\text{Nº de hombres}}{\text{Nº de mujeres}}$

Indica cuántos hombres hay por cada mujer. Si se amplifica por 100, se sabrá cuantos hombres hay por cada 100 mujeres, en Chile 1982 había 96 hombres por cada 100 mujeres.

$$\text{Chile 1982} = \frac{5.521.067}{5.754.373} \times 100 = 95.9$$

##### Otro ejemplo :

En el programa de atención maternal se desea comparar la relación entre controles y consultas de morbilidad otorgadas en dos Servicios de Salud en 1982.

Servicio de Salud	Atención maternal	
	Controles	Consultas morbilidad
Oriente	72.154	72.568
Sur	72.029	87.041

El examen de estas cifras absolutas hace un poco difícil la comparación. En una forma gruesa se puede decir que ambos Servicios dieron igual número de controles y que, en cambio, el número de consultas por morbilidad fue muy superior en el Servicio Sur. Resulta más clara la comparación si se calculan los cuocientos entre el número de controles y el número de consultas en cada uno de los Servicios.

Servicio Oriente  $72.154/72.568 = 1$  control por cada consulta.

Servicio Sur  $72.029/87.041 = 0.8$  controles por cada consulta.

Se establece que el Servicio Oriente ha dado más controles por consulta que el Servicio Sur.

## 2. Proporciones

Son cuocientos entre dos cantidades de igual naturaleza. Describen la fracción que una serie de sucesos que figuran en el numerador representa con respecto al total de sucesos de igual índole. Cuando el resultado de este cuociente se multiplica por 100 resulta un porcentaje, que es la forma habitual de calcular esta frecuencia relativa.

Ejemplo: En Chile en 1982 el Sistema Nacional de Servicios de Salud controló el estado nutricional de 1.160.813 niños menores de 6 años. En el mismo año la Región Metropolitana controló 390.464 niños de igual edad. Como la Región Metropolitana es una parte del Sistema Nacional se puede calcular el porcentaje que representan los controles de esta Región con respecto al total del país:

$$\frac{390.464}{1.160.813} \times 100$$

De este modo se sabe que 34% del total de niños menores de seis años en control nutricional en el país, pertenecen a la Región Metropolitana.

Es importante insistir que tanto los hechos que figuran en el numerador como los del denominador deben ser de igual naturaleza. De este modo el resultado expresa la importancia relativa que el dato del numerador tiene con respecto al total.

Los porcentajes tienen la ventaja de permitir una comparación fácil de series que tienen totales diferentes, al referirlos a una base común que en esto caso es 100. Si suponemos dos Provincias en que se desea conocer si la mortalidad del menor de 28 días es diferente en importancia con respecto al total de niños menores de 1 año, es más sencillo calcular los porcentajes que representan las defunciones de menores de 28 días con respecto al total de defunciones de menores de 1 año.

Provincia	Defunciones de menores de 1 año	Defunciones de menores de 28 días	
		Nº	%
Concepción	502	242	48,2
Bío-Bío	287	131	45,6

En la provincia de Concepción las defunciones de menores de 28 días representan el 48,2% del total de defunciones infantiles, en cambio en la provincia de Bío-Bío representan el 45,6%.

### Limitaciones de los porcentajes y necesidad del cálculo de tasas

A pesar de su utilidad, los porcentajes tienen limitaciones. Si se estudian, por ejemplo, las muertes por accidentes en dos grupos de edades en un país X nos encontramos con lo siguiente:

Grupo de edad	Total de defunciones	Defunciones por accidente	
		Nº	%
15 a 24	48.999	12.763	26,0
65 a 74	306.025	11.425	3,7

En este caso podría concluirse que los accidentes son un peligro más serio para los jóvenes, en los que más de una cuarta parte de las defunciones se debe a accidentes, que para las personas de mayor edad, en las que los accidentes causan menos del 4% de las defunciones.

Las cifras anteriores no expresan realmente el riesgo de morir por accidente, sino la importancia relativa que esta causa tiene en el total de defunciones de cada grupo de edad. El conocimiento del riesgo no se obtiene con el cálculo de los porcentajes; para ello hay que introducir en la comparación un elemento importante que es la población expuesta al riesgo de sufrir accidentes. El resultado que se obtiene al dividir el número de muertes debidas a accidentes por la población expuesta al riesgo de sufrir un accidente es lo que se denomina tasa de mortalidad por accidente.

### 3. Tasas

Una tasa es un cociente formado por tres elementos :

- Un numerador, que consiste en el número de veces que ocurrió un determinado hecho en un período de tiempo dado y en un área determinada.  
Por ejemplo, el número de casos de una enfermedad que se registró en un área durante un año.
- Un denominador, que es la población expuesta al riesgo de que le suceda el fenómeno que aparece en el numerador.
- Una constante por la cual se multiplica el cociente. Debido a que el cociente resultante en una tasa es siempre de valor inferior a la unidad, éste se multiplica por 100, 1.000, 10.000 ó 100.000 de modo de tener cifras superiores a la unidad, lo que facilita la interpretación. En efecto, es más fácil entender que la tasa de mortalidad de una región es 8 por 1.000 habitantes que decir que es 0,008 por habitante.

### Requisitos generales de las tasas

Es necesario que en una tasa haya concordancia entre el numerador y el denominador en tres aspectos importantes: la naturaleza del hecho, la zona geográfica y el período de tiempo dentro del cual ocurre el hecho.

En relación con la naturaleza del hecho, debe usarse en el denominador la población de la cual haya emanado el hecho del numerador. Así, no podríamos tener una tasa de mortalidad por cáncer de la próstata si en el denominador figura la población femenina.



El área geográfica debe ser la misma para el numerador que para el denominador.

Con respecto al tiempo, las tasas se calculan generalmente sobre una base anual. Se presenta un problema en cuanto al denominador de la tasa, ya que debido a que la población varía a lo largo del año, pueden hacerse distintas estimaciones de ella. Si la población se estima al comienzo del período no representa toda la población expuesta ya que en esta población no figuran por ejemplo, los niños que nacerán durante el año. Si la población que se usa es la estimada al final del año sucede lo contrario, ya que no aparecerán en ella los que han fallecido y los que han emigrado en el curso del año.

De aquí que es de uso habitual como representativa de la población media expuesta al riesgo la estimación a mitad del período, es decir, al 30 de junio del año en estudio.

### Tipos de tasas

En general pueden distinguirse dos tipos principales de tasas:

- a) tasas crudas o brutas.
- b) tasas específicas.

Quando en el denominador figura el total de la población se habla de tasas crudas porque no se consideran características como edad, sexo, etc. Es una medición gruesa de la fuerza de ocurrencia de un hecho.

Quando en el denominador se usa sólo cierto sector de la población, por ejemplo, la población de 20 a 25 años (en el numerador debe figurar el hecho referido que afecta sólo a este grupo de edad) se habla de tasas específicas. Estas tasas son más refinadas y miden con mayor exactitud el riesgo que se desea conocer, ya que en general los riesgos son diferentes según las características de las personas.

Por ejemplo, la mortalidad es muy diferente en algunos grupos de edad y la tasa cruda es sólo una especie de promedio de las diferentes tasas específicas. A veces se habla impropriamente de que una tasa es específica. Tal es el caso de la tasa de mortalidad por una causa determinada, por ejemplo tuberculosis. Si en el numerador figuran todas las defunciones por tuberculosis en el denominador debe estar toda la población y es por lo tanto una tasa cruda por una causa específica.

Las tasas que habitualmente se usan en Salud Pública se refieren a la mortalidad, la morbilidad, la letalidad y la fecundidad.

#### 3.1. Tasa bruta de mortalidad

Su numerador incluye la totalidad de las defunciones de ambos sexos, de todas las edades y por todas las causas, registradas a lo largo de un año calendario en un área determinada. Su denominador es la población total de esa misma área estimada a mitad de período, es decir, al 30 de junio del mismo año. Tal como ocurre con todas las tasas de mortalidad, debido a que en la población expuesta al riesgo de morir sólo algunos individuos han muerto en el término del año calendario, el denominador es siempre mayor que el numerador y, para obtener cifras enteras es necesario amplificar el cociente entre defunciones y población por una constante que, en el caso de la tasa bruta es 1.000.

Tasa bruta de mortalidad =

$$= \frac{\text{Nr total de defunciones en un área y año determinados}}{\text{Población total del área al 30 de junio de ese año}} \times 1.000$$

### Según causa

Tasa de mortalidad por causa =

$$= \frac{\text{Defunciones por una causa en un área y año determinados}}{\text{Población total al 30/junio de ese año y área}} \times 100.000$$

El denominador de las tasas por causa, en general, es la población total y por consiguiente se trata de tasas crudas por una causa o grupo de causas específicas.

La construcción de estas tasas implica separar el conjunto de todas las muertes diversos sub conjuntos atendiendo a la causa de muerte. Dichas muertes, si no hay otra especificación adicional, incluyen las defunciones de cualquier edad y ambos sexos que han ocurrido por una misma causa o grupo de causas.

Debido a la necesidad de disponer de tasas por causas de muerte cuya magnitud en la población pueda ser muy pequeña y a fin de que la magnitud de las tasas de mortalidad por las diferentes causas sea fácilmente comparable, la constante que en ellas se utiliza es 100.000.

Tasa de mortalidad materna =

$$= \frac{\text{Muertes debidas a complicaciones del embarazo, parto o puerperio}}{\text{Nacidos vivos en ese año y área}} \times 1.000 \text{ (ó } \times 10.000 \text{)}$$

Se denominan muertes maternas aquellas cuya causa está relacionada con complicaciones del embarazo, parto o puerperio y ellas constituyen el numerador de la tasa.

Su denominador podrían ser las mujeres entre 15 y 49 años pero el riesgo específico que indica el numerador sólo afecta a aquellas que en dicho año han tenido un embarazo, por lo tanto lo más adecuado sería colocar el número de embarazos. Como habitualmente no se dispone de información fidedigna respecto a este dato, se ha convenido internacionalmente utilizar como denominador el número de nacidos vivos del mismo año en que sucedieron las muertes del numerador.

La tasa de mortalidad materna se define como la relación entre el número de defunciones por causas relacionadas con las complicaciones del embarazo, parto o puerperio ocurridas en un año y área dadas y el número de nacidos vivos en el mismo año y área. Se puede expresar por 1.000 ó por 10.000.

### 3.2. Tasas específicas de mortalidad

#### Según sexo:

El riesgo de morir difiere según el sexo. Por ello es conveniente medir por separado la mortalidad de hombres y de mujeres.

Tasa Mortalidad Masculina =

$$= \frac{\text{Defunciones masculinas en un área y año determinados}}{\text{Población masculina al 30/VI de ese año y área}} \times 1.000$$

Tasa Mortalidad Femenina =

$$= \frac{\text{Defunciones femeninas en un área y año determinados}}{\text{Población femenina al 30/VI de ese año y área}} \times 1.000$$

Igual que la tasa bruta de mortalidad, ambas tasas se amplifican por 1.000.

Debido a que sus denominadores son diferentes estas dos tasas no se pueden sumar directamente para reconstruir la tasa bruta de mortalidad.

### Según edad :

La mortalidad difiere marcadamente según la edad. Por eso corrientemente la medición de la mortalidad requiere medir el riesgo de muerte por edades. Al elaborar las tasas de mortalidad por edad puede llegarse a tal grado de especificación que los subconjuntos de defunciones incluyan sólo edades simples, es decir, se elabore una tasa para cada año de edad. Sin embargo, lo habitual es que se trabaje con grupos de edades, usándose frecuentemente grupos quinquenales de edad o bien grupos de mayor amplitud. Sólo para las edades más jóvenes, en que el riesgo de morir cambia más rápidamente con la edad, está justificado construir tasas de mortalidad por edades simples o aún por intervalos que sean menos amplios que 1 año.

Tasa de mortalidad por edad =

$$= \frac{\text{Defunciones de un grupo de edad en un área y año determinados}}{\text{Población de ese grupo de edad al 30/VI de ese año y área}} \times 1.000$$

Todas las tasas de mortalidad por edad se amplifican por 1.000. Estas tasas se pueden calcular separadamente para cada sexo. En tales casos la doble especificación de sexo y edad debe hacerse tanto para las defunciones como para la población. Ejemplo:

Tasa mortalidad masculina de 20 - 24 años =

$$= \frac{\text{Defunciones masculinas de 20-24 en un área y año determinados}}{\text{Población masculina de 20-24 años al 30/VI para ese año y área}} \times 1.000$$

Un caso especial dentro de las tasas de mortalidad por edad lo constituyen las muertes de los menores de un año. El riesgo de morir es considerablemente más alto en el primer año de vida que en las edades siguientes, salvo las edades muy avanzadas.

Es precisamente en esta edad cuando la mortalidad es más sensible a los efectos del ambiente y si las tasas son altas una buena proporción de estas defunciones son evitables. Por ello esta medida es un indicador usual del nivel de salud e interesa particularmente conocerla.

Tasa de mortalidad infantil =

$$= \frac{\text{Defunciones de niños menores de 1 año en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Tal como en la tasa bruta de mortalidad y las tasas de mortalidad por sexo y edad, la constante que se utiliza en esta tasa es 1.000.

El numerador de la tasa de mortalidad infantil incluye las defunciones de ambos sexos y por todas las causas que ocurren dentro de un año calendario y en un área determinada en los niños que aún no han cumplido su primer año de vida. Dada la naturaleza de su numerador la tasa de mortalidad infantil tiene el carácter de una tasa de mortalidad por edad. Por lo tanto, debería esperarse que su denominador fuera la población de menores de 1 año de edad, estimada a mitad del mismo año calendario a que se refieren las muertes. Sin embargo, hay razones metodológicas por las cuales se hace necesario el uso de otro denominador.

Entre estas razones está el hecho de que la población menor de 1 año se omite en los censos en una proporción mayor que la de cualquiera otra edad, y por ello su tamaño, para un año censal y con mayor razón en las estimaciones para los años posteriores al censo, son más inexactas que para los grupos de edades mayores. Por otra parte, los niños menores de 1 año que existen en una población depende del nivel y las tendencias de la natalidad en los años recientes. En cambio, en los grupos de edades mayores los efectivos de población son menos sensibles a las modificaciones de la natalidad en los años inmediatamente precedentes.

Es por esto, que para estar a cubierto de las variaciones que existen entre los países respecto a la cobalidad de los censos y de las fluctuaciones que puede experimentar el nivel de la natalidad, se ha convenido internacionalmente en utilizar como denominador de la tasa de mortalidad infantil la cifra de nacidos vivos del año, en lugar de la población estimada de menores de 1 año.

- La tasa de mortalidad infantil se subdivide en dos componentes:

Tasa de mortalidad neonatal =

$$= \frac{\text{Defunciones niños menores de 28 ds. en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Esta tasa mide la frecuencia de muertes que ocurren en los menores de 28 ds. en un año calendario y en un área determinada por cada 1.000 nacidos vivos en ese mismo año y área.

Tasa de mortalidad infantil tardía =

$$= \frac{\text{Defunciones de niños de 28 ds. a 11 meses en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

La tasa de mortalidad infantil tardía mide la frecuencia de muertes que ocurren en el primer año de vida a partir del 28º día, en un año calendario y área dada por 1.000 nacidos vivos en ese año y área.

3. Así como entre las muertes del primer año es conveniente distinguir las que ocurren en las primeras 4 semanas del resto de las muertes infantiles, también es útil analizar separadamente las muertes de la primera semana de vida de las correspondientes a las 3 semanas siguientes. Si se refieren estos nuevos dos subconjuntos a la misma cifra de nacidos vivos del año se obtienen dos nuevas tasas que sumadas equivalen a la tasa de mortalidad neonatal. Ambas se expresan igualmente por 1.000. La tasa de mortalidad de la primera semana se denomina tasa de mortalidad neonatal precoz y la de la segunda a cuarta semana tasa de mortalidad neonatal tardía.

Tasa de mortalidad neonatal precoz =

$$= \frac{\text{Defunciones de menores de 7 días en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Esta tasa mide la frecuencia de muertes que ocurren en la primera semana de vida en un año calendario y área dada por cada 1.000 nacidos vivos del mismo año y área.

Tasa de mortalidad neonatal tardía =

$$= \frac{\text{Defunciones de niños de 7 a 27 días en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Mide la frecuencia de muertes que ocurren entre la segunda y cuarta semana de vida en un año calendario y área dada por cada 1.000 nacidos vivos del mismo año y área.

Tasa de mortalidad fetal tardía (o mortalidad) =

$$= \frac{\text{Defunciones fetales tardías (28 y + semanas de gestación) en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Según el momento de la gestación en que se produce la muerte del producto de la concepción, las defunciones fetales se clasifican en precoces (menos de 20 semanas de gestación) intermedias (20 a 27 semanas) y tardías (28 y más semanas de gestación). Las defunciones fetales tardías corresponden a los mortinatos y las precoces e intermedias a los abortos.

El registro de las defunciones fetales tiene una omisión importante. Esta omisión afecta principalmente a las defunciones fetales precoces.

Para las defunciones fetales tardías en cambio, el registro proporciona una información más completa, aunque siempre subestima la magnitud real del problema. Su denominador también son los nacidos vivos por las razones expuestas en la tasa de mortalidad materna.

Tasa de mortalidad perinatal =

$$= \frac{\text{Defunciones fetales tardías} + \text{defunciones de niños menores de 7 días en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1,000$$

Esta tasa mide el riesgo de muerte que implica para el producto de la concepción el paso de la vida intrauterina a la vida extrauterina.

### 3.3 Medición de la morbilidad

El estudio de la morbilidad tiene serias dificultades. Desde luego, a diferencia de la muerte que ocurre una sola vez y en un momento bien definido y es un hecho permanente, la enfermedad puede ocurrir varias veces en la vida de un individuo, ya que se trata de una misma enfermedad o de enfermedades distintas y por último ellas pueden tener duración variable.

En lo que se refiere a la medición de la enfermedad se pueden distinguir tres tipos de unidades:

1) personas enfermas, 2) enfermedades, 3) episodios de enfermedad.

Por ejemplo, si una persona tiene durante el año 2 resfrios y 3 episodios diarreicos, se contabilizará:

a) persona enferma; b) 2 enfermedades; c) 5 episodios.

Por este motivo el Comité de Expertos en Estadísticas de Salud recomienda que en las estadísticas de morbilidad se especifique claramente a cual de estos tres criterios se refieren los datos.

En la medición de la morbilidad interesa fundamentalmente medir la frecuencia de la enfermedad en la población, su duración y su gravedad.

#### 3.3.a. Medición de la frecuencia de la enfermedad

Se distinguen dos tipos: la incidencia y la prevalencia.

##### Tasa de incidencia

Se denomina incidencia al número de casos nuevos que se presentan en un período de tiempo. Se refiere a enfermedades que comienzan durante un período definido y la tasa mide la frecuencia de acontecimientos que ocurren durante el período. En la tasa de incidencia se incluyen en el numerador los casos nuevos (enfermedades o enfermos) registrados durante el período y el denominador se refiere a la población estimada en el punto medio del período. Las tasas de incidencia pueden ser anuales pero también pueden referirse a cualquiera otra unidad de tiempo.

Tasa de incidencia =

$$= \frac{\text{Número de casos nuevos en el período}}{\text{Población a mitad del período}} \times 100.000$$

La tasa de incidencia muestra la dinámica de la enfermedad y expresa el riesgo de enfermar que tiene la población durante el período observado.

-Tasa de prevalencia

Prevalencia : es el número de casos (nuevos y antiguos) que se registran en un tiempo o momento dado, por ejemplo, el primer día de un mes o el último día de un año o el promedio diario dentro de un período de tiempo. La tasa de prevalencia tiene como numerador el número de casos que están presentes en ese momento y como denominador la población estimada para el mismo momento.

Tasa de prevalencia =

$$= \frac{\text{Número de casos existentes en un momento dado}}{\text{Población en ese momento}} \times 100.000$$

La tasa de prevalencia es una medida relativa cuyo sentido es comparable a la información que proporcionan los censos de población y mide sólo lo que existe o prevalece en ese momento. Es necesario hacer notar que en el numerador figuran todos los casos tanto los que se iniciaron antes del momento de medición como los casos nuevos que aparecen en ese momento.

Tratándose de enfermedades crónicas la prevalencia refleja mejor que la incidencia la magnitud del problema en la comunidad.

### 3.3.b. Medición de la gravedad de la enfermedad

Un aspecto de la morbilidad cuyo conocimiento tiene gran interés es la gravedad de la enfermedad. Ella puede medirse en términos de la incapacidad que produce. Por ejemplo, una enfermedad menor es aquella que no es causa de ausencia del trabajo. Esto hace necesario tener una escala de incapacidad para medir la severidad del cuadro. Además la medición tiene el problema de que la gravedad depende no sólo de la enfermedad sino que también de las características de los individuos que la padecen. Por ejemplo, un resfriado común puede ser motivo para que una persona guarde cama, mientras otro individuo con un resfriado de iguales condiciones continúa desarrollando sus actividades.

Por estas dificultades el índice de gravedad de una enfermedad que más se utiliza es la tasa de letalidad, que establece la relación entre los fallecidos por enfermedad y los enfermos que padecen esa enfermedad.

Tasa de letalidad =

$$= \frac{\text{Número de defunciones por una enfermedad dada}}{\text{Número de enfermos de esa enfermedad}} \times 100$$

Mide la frecuencia con que se produce la muerte en una enfermedad. Esto es la tasa que permite establecer el pronóstico de las enfermedades.

### 3.3.c. Medición de la duración de la enfermedad

La duración de la enfermedad es un dato que interesa medir, entre otras razones, porque la enfermedad de mayor duración significa mayor costo. Puede hacerse esta medición en forma de un promedio. Por ejemplo, 60 enfermos de tifoides estuvieron en cama un total de 1.080 días, la duración de la enfermedad es entonces:

$$\text{Duración} = \frac{1.080}{60} = 18 \text{ días en promedio}$$

Para la medición de la duración es necesario definir previamente que se entiende por enfermedad. En este caso la duración se refiere al tiempo promedio de estadía en cama de los enfermos. Otras definiciones podrían tomarse en cuenta, por ejemplo, el día de los primeros síntomas o el día en que se hizo el diagnóstico, etc.

El promedio puede obtenerse no solo en relación a los enfermos (60 en el ejemplo anterior) sino que puede obtenerse para episodios de enfermedad. Por ejemplo: en una escuela se registraron los resfriados de los alumnos y se tuvo un total de 100 resfriados en el año. La duración total de los resfriados fue de 500 días. La duración media de cada episodio fue, por lo tanto de 5 días.

### 3.4. Medición de la fecundidad

La medición de la fecundidad se hace a través de diferentes tipos de tasas que tratan de medir los niveles del fenómeno en un área.

#### Tasa bruta de natalidad

Es una tasa simple que relaciona los nacidos vivos registrados en un área geográfica durante un año con la población total de esta área.

Tasa bruta de natalidad =

$$= \frac{\text{Nacidos vivos en un área y año determinados}}{\text{Población total al 30/VI en ese año y área}} \times 1.000$$

Como incluye a la población total (de todas las edades y de ambos sexos) no puede interpretarse como una probabilidad porque en el denominador hay población que no está expuesta al riesgo de tener un niño. Expresa más bien la frecuencia de los nacimientos por cada 1.000 habitantes.

Las tasas de natalidad son prácticamente las únicas medidas de fecundidad que es posible calcular para áreas geográficas pequeñas y permite estudiar las tendencias del fenómeno en un área determinada.

Quando se comparan áreas diferentes hay que ser extremadamente cuidadoso en la interpretación porque puede haber diferencias en la estructura de la población especialmente en lo que se refiere a la composición por edad de la población femenina y esta diferencia puede por sí sola determinar diferencias en las tasas de natalidad.

#### Tasa de fecundidad general

Es esta una tasa más específica ya que tiene un denominador la población potencialmente expuesta al riesgo de tener un nacido vivo: la población femenina en edad fértil.

#### Tasa de fecundidad general =

$$= \frac{\text{Nacidos vivos en un área y año determinados}}{\text{Población femenina de 15 a 49 años al 30/VI}} \times 1.000$$

en ese año y área.

Al tomar en cuenta solamente a las mujeres y en el grupo de edad expuesto al riesgo es una tasa más útil para hacer comparaciones entre zonas o comparaciones internacionales.

#### Tasa de fecundidad por edad

Esta tasa tiene un nuevo refinamiento y es más específica ya que toma en cuenta no solo el sexo, sino la composición por edad. En efecto, en su numerador se anotan los nacimientos de madres de un grupo de edad determinada y en el denominador la población femenina de esa edad.

#### Tasa de fecundidad por edad =

$$= \frac{\text{Nacidos vivos de mujeres de un grupo de edad en un área y año determinados}}{\text{Población femenina de ese grupo de edad al 30/VI}} \times 1.000$$

en ese año y área.

Ej.: 
$$\frac{\text{Nacidos vivos de mujeres de 15 a 19 años en un área y año determinados}}{\text{Población femenina de 15 a 19 años al 30/VI}} \times 1.000$$
 en ese año y área.

Por lo general las tasas de fecundidad por edad se calculan para grupos quinquenales de edades comprendidas entre los 15 y los 49 años, es decir, se calculan 7 tasas de fecundidad por edad.

#### Otras medidas de fecundidad

Los estudios demográficos más finos de la fecundidad utilizan además de las tasas anteriores, las llamadas tasas de reproducción que tratan de medir el aporte futuro de la fecundidad al reemplazo de la población haciendo una corrección en los nacimientos utilizando la proporción de nacimientos femeninos.

Como se trata de tasas usadas por especialistas remitimos al lector a los textos de Demografía para su estudio.



## Capítulo IV

### INTERPOLACION DE FUNCIONES

#### POLINOMIOS DE INTERPOLACION

El problema más simple de interpolación consiste en lo siguiente: Dado un intervalo  $[a, b]$  en el cual se hallan especificados los puntos  $x_0, x_1, x_2, \dots, x_n$ , y los valores de una cierta función ;  $f(x_0) = y_0, f(x_1) = y_1, f(x_2) = y_2, \dots, f(x_n) = y_n$ , se requiere construir una función  $P(x)$  llamada función de interpolación que pertenezca a una clase conocida de funciones (en particular la clase de las funciones polinomiales ) y que tome los mismos valores en los puntos de interpolación que la función  $f(x)$ , es decir, tal que:

$$(1) \quad P(x_0) = y_0; \quad P(x_1) = y_1, \quad P(x_2) = y_2, \dots, \quad P(x_n) = y_n$$

Geoméricamente esto significa que debemos hallar un polinomio  $y=P_n(x)$  de grado no superior a "n" que pase por los PUNTOS DE INTERPOLACION y que satisfaga la condición (1). Conviene señalar que en nuestro caso la función  $y=f(x)$  estará siempre definida mediante una tabla de datos resultado de alguna observación.

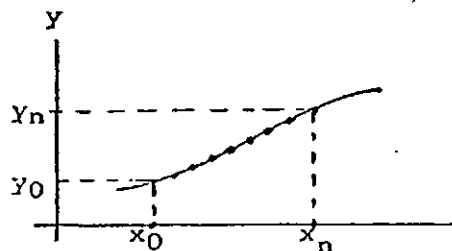


fig. (1)

El polinomio de interpolación encontrado se utiliza ordinariamente para aproximar los valores de la función  $y = f(x)$  para valores del argumento "x" que no sean los puntos de interpolación  $x_0, x_1, x_2, \dots, x_n$ . Esta operación se llama INTERPOLACION DE LA FUNCION. En general se distingue entre INTERPOLACION en sentido estricto cuando "x" pertenece al intervalo  $[x_0, x_n]$ , es decir, cuando el valor de "x" es intermedio entre  $x_0$  y  $x_n$  y EXTRAPOLACION cuando el valor de "x" no pertenece al intervalo  $[x_0, x_n]$ . Sin embargo nos referiremos siempre al término interpolación para significar tanto la interpolación en sentido estricto como la extrapolación.

EJEMPLO. El censo realizado, cada cinco años, en una ciudad, arroja los siguientes datos para los últimos 15 años.

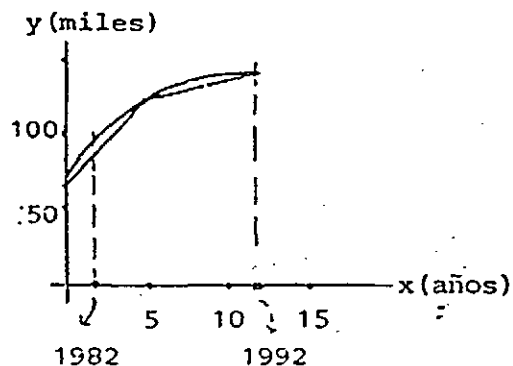
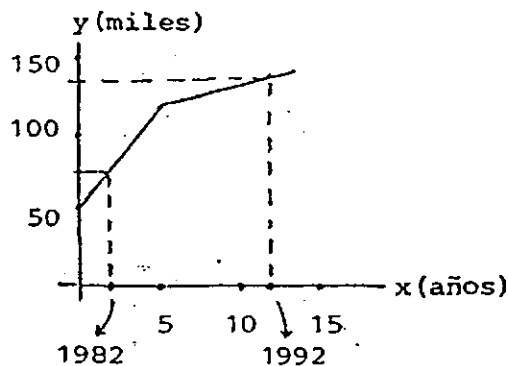
x	1980	1985	1990
y	70.350	120.458	130.790

Se pide determinar aproximadamente la cantidad de habitantes que había en el año 1982 y los que habrá en el año 1992.

SOLUCION. Este problema puede tener muchas soluciones, dependiendo del polinomio de interpolación que escojamos. Un gráfico de los datos es el siguiente: (t = 0 en 1980).

fig.(2)

fig.(3)



En la figura 2 hemos unido los puntos de interpolación con una línea poligonal continua suponiendo que el comportamiento del crecimiento de la población podría ser modelado por una función lineal (polinomio de primer grado) cada cinco años. Sin embargo podemos suponer, también, que el comportamiento del crecimiento de la población puede estar modelado por una función cuadrática, esto es, por un polinomio de segundo grado, cuyo gráfico es una parábola. Recordemos que una parábola queda completamente determinada por tres de sus puntos. ¿Cuál de las dos funciones es la que nos proporcionará la información más ajustada a la realidad?. La elección del grado del POLINOMIO INTERPOLANTE no es fácil y con frecuencia está ligado a la posesión de información adicional por parte del investigador que está interpolando.

En realidad es posible hallar la ecuación de una recta y de una parábola sin recurrir a los métodos, algo sofisticados, de interpolación; para esto sería suficiente resolver un sistema de

ecuaciones lineales en dos y tres variables respectivamente. Sin embargo cuando se tiene una tabla de datos, que es lo usual en demografía, resulta, a la postre, más fácil y más cómodo este tratamiento especial mediante los llamados métodos de interpolación.

PRIMERA FORMULA DE INTERPOLACION DE NEWTON  
 VALORES DEL ARGUMENTO IGUALMENTE ESPACIADOS

Supongamos que se tiene una función dada mediante una tabla como la siguiente:

x	$x_0$	$x_1$	$x_2$	.....	$x_n$
Y	$Y_0$	$Y_1$	$Y_2$	.....	$Y_n$

para valores igualmente espaciados de la variable independiente, esto es,  $x_i = x_0 + ih$  ( $i=0,1,2,\dots,n$ ), donde "h" (espaciado) es la distancia entre cada dos puntos. Se requiere hallar un polinomio  $P_n(x)$  de grado no superior a "n" tal que:

$$P_n(x_0) = Y_0, \quad P_n(x_1) = Y_1, \quad \dots, \quad P_n(x_n) = Y_n$$

Esto significa, por ejemplo, que dados dos puntos se hallará un polinomio de primer grado, dados tres puntos el polinomio será de segundo grado, dados cuatro puntos el polinomio será de tercer grado, etc. La fórmula de interpolación de Newton, que nos da el polinomio de interpolación, se escribe ordinariamente en la forma siguiente:

$$(2) \quad P_n(x) = y_0 + q\sigma y_0 + \frac{q(q-1)}{2!} \sigma^2 y_0 + \frac{q(q-1)(q-2)}{3!} \sigma^3 y_0 + \dots$$

$$+ \frac{q(q-1)(q-2)\dots(q-n+1)}{n!} \sigma^n y_0, \quad \text{donde } q = \frac{x-x_0}{h} \text{ y las}$$

$\sigma^i y_0$  son las DIFERENCIAS FINITAS construidas utilizando los puntos dados por la tabla de datos y que están calculados en la TABLA DE DIFERENCIAS HORIZONTAL.

Esta fórmula resulta muy conveniente para interpolar una función dada cerca del punto inicial  $x_0$ . Conviene indicar que cualesquiera de los valores de "x" dado por la tabla puede ser considerado como valor inicial  $x_0$ ; dependiendo del "lugar" donde se quiera interpolar.

#### INTERPOLACION LINEAL

Si la tabla de datos está conformada solamente por dos puntos el polinomio de interpolación será de primer grado; en cuyo caso tendremos la ecuación de una recta. Haciendo  $n = 1$  en la expresión (2) resulta:

$$P_1(x) = y_0 + q\sigma y_0$$

que es la fórmula de interpolación lineal.

#### INTERPOLACION CUADRATICA

Si la tabla de datos está formada por tres puntos, entonces, el polinomio de interpolación será de segundo grado; en cuyo caso tendremos la ecuación de una parábola. Haciendo  $n = 2$  en la expresión (2) resulta:

$$P_2(x) = y_0 + q\sigma y_0 + \frac{q(q-1)}{2i} \sigma^2 y_0$$

que es la fórmula de interpolación cuadrática.

Podemos seguir obteniendo los polinomios de tercer grado, cuarto grado, etc, sin embargo es suficiente para nuestros propósitos trabajar con los polinomios de primer y segundo grado.

Estamos ahora en condiciones de resolver completamente el problema planteado al principio, esto es, hallar aproximadamente la cantidad de habitantes que había en el año 1982 y que habrá en el año 1992.

CASO LINEAL

La tabla de diferencias horizontal de datos es la siguiente:

	x	y	$\sigma y_0$	$\sigma^2 y_0$
1980	0	70.350	50.108	-40.566
1985	5	120.458	9.542	
1990	10	130.000		
1992	12			

Si usamos interpolación lineal para hallar la población en el año 1982, esto es, cuando  $x = 2$ , tenemos que usar  $x_0 = 0$  como valor inicial y considerar solamente las diferencias hasta  $\sigma y_0$ , tal como indica la fórmula. De esto se desprende entonces que:

$$x_0 = 0, y_0 = 70.108, \sigma y_0 = 50.108 \text{ y } h = 5$$

Por otro lado, puesto que, el punto en el cual queremos interpolar es  $x = 2$  resulta que:

$$q = \frac{2-x_0}{5} = \frac{2-0}{5} = \frac{2}{5} = 0,4$$

Finalmente de  $P_1(x) = y_0 + q\sigma y_0$ , resulta:

$$P_1(2) = 70.108 + 0,4(50.108)$$

$$P_1(2) = 90.151 \text{ personas.}$$

Por lo tanto en el año 1982 habría habido aproximadamente 90.151 personas.

Para responder a la pregunta ¿ Cuántos habitantes habrá en el año 1992 ? tenemos que tomar la recta formada por los dos últimos puntos puesto que, nuestra respuesta, está en la PROLONGACION de la recta formada por ellos. En esta situación nuestro valor inicial será  $x_0=5$ . Observe que estamos tratando de interpolar en un punto más cercano a 10 que a 5, sin embargo no podemos tomar como valor inicial a 10 porque en ese caso no tendríamos recta.

En efecto:

$$x_0 = 5, y_0 = 120.458, \sigma y_0 = 9.542, h = 5$$

Como queremos interpolar en  $x = 12$  resulta que:

$$q = \frac{12-x_0}{5} = \frac{12-5}{5} = \frac{7}{5} = 1,4$$

Luego de  $P_1(x) = y_0 + q\sigma y_0$  tenemos:

$$P_1(12) = 120.458 + 1,4(9.542)$$

$$P_1(12) = 133.458 \text{ personas.}$$

Por lo tanto en el año 1992, podría haber alrededor de 133.458 personas.

#### CASO CUADRATICO

Utilizando la misma tabla de diferencias horizontal debemos interpolar en  $x = 2$  considerando:

$$x_0 = 0, y_0 = 70.108, \sigma y_0 = 50.108, \sigma^2 y_0 = -40.566 \text{ y } h = 5$$

Puesto que el valor de  $q = 0,4$ , tal como lo habíamos calculado antes, resulta de:

$$P_2(x) = y_0 + q\sigma y_0 + \frac{q(q-1)}{2} \sigma^2 y_0, \text{ que}$$

$$P_2(2) = 70.350 + 0,4(50.108) + \frac{0,4(0,4-1)}{2} (-40.566)$$

$$P_2(2) = 90.393,2 + (-0,12)(-40.566) = 95.261,1 \text{ personas.}$$

Por lo tanto en el año 1982 habría habido aproximadamente 95.261 personas. Este resultado superior en alrededor de 5.000 individuos, en relación al caso lineal, no debería asombrarnos porque el segmento de parábola, entre los dos primeros puntos, está sobre el segmento de recta de dichos puntos. Ver fig(3).

¿ Cuántos habitantes habrá en el año 1992 ?. Puesto que la parábola necesita tres puntos para quedar determinada tenemos que usar nuevamente el mismo polinomio para interpolar en  $x = 12$ , con  $x_0 = 0$ , pero con diferente "q". En efecto:

$$q = \frac{x-x_0}{h} = \frac{12-0}{2} = 6$$

Resulta entonces que:

$$P_2(12) = 70.350 + 6(50.108) + \frac{12(12-1)}{2} (-40.566)$$

$$P_2(12) = 370.998 + 66(-40.566) = 103.262,4 \text{ personas.}$$

Observemos que esta cantidad, menor que la del año 1990 se debe al hecho que alrededor del punto (10, 130.790) la parábola empieza a decrecer ya que la tendencia de la población sigue ese comportamiento. Sin embargo es poco probable que, en la realidad, la población disminuya tan rápidamente. Es en estos casos cuando la interpolación se transforma en un arte puesto que el investigador debe utilizar información adicional sobre el fenómeno que está estudiando.

EJEMPLO. La siguiente es una tabla de los sobrevivientes a edades exactas, de Guatemala del año 1950, comprendida entre los 20 y 50 años.

x	20	30	40	50
$l_x$	64.331	59.036	52.820	44.739.

Hallar a)  $l_{22}$  y b)  $l_{45}$  utilizando interpolación lineal y cuadrática respectivamente.

SOLUCION. Construiremos en primer lugar la tabla horizontal de diferencias finitas.

x	$l_x$	$\sigma y_0$	$\sigma^2 y_0$	$\sigma^3 y_0$
20	64.331	-5.296	-920	-945
30	59.036	-6.216	-1.865	
40	52.820	-8.081		
50	44.739			

a) Puesto que  $l_{22}$  está entre los 20 y los 30 años el valor inicial  $x_0=20$ . Tenemos, entonces lo siguiente:

$$x_0 = 20, y_0 = 64.331, \sigma y_0 = -5.296, h = 10, q = \frac{x-20}{10} = \frac{22-20}{10} = 0,2.$$

Aplicando la fórmula de interpolación lineal resulta:

$$\begin{aligned} l_{22} &= y_0 + q\sigma y_0 = 64.331 + 0,2(-5.296) = \\ &= 64.331 - 1.059,2 = 63.271,8 \end{aligned}$$

Por lo tanto hay, a la edad de 22 años, 63.272 personas aproximadamente.

b) Como  $l_{45}$  está entre los últimos tres puntos debemos aplicar la interpolación cuadrática entre ellos. En este caso  $x_0 = 30$ . Resulta entonces que:

$$x_0=30, y_0=59.036, \sigma y_0=-6.216, \sigma^2 y_0=-1.865, h=10 \text{ y } q = \frac{45-30}{10} = 1,5$$

Aplicando la fórmula de interpolación cuadrática tenemos:

$$\begin{aligned} l_{45} &= 59.036 + 1,5(-6.215) + \frac{1,5(1,5-1)}{2}(-1.865) \\ l_{45} &= 59.358,5 - 9.3322,5 + 669,4 = 50.382,9 \end{aligned}$$

Por lo tanto hay, a la edad de 45 años, 50.383 personas aproximadamente.

#### SEGUNDA FORMULA DE INTERPOLACION DE NEWTON VALORES DEL ARGUMENTO DESIGUALMENTE ESPACIADOS

¿ Qué ocurre cuando los valores de la variable independiente están desigualmente espaciados ?. En este caso la primera fórmula de interpolación de Newton no puede aplicarse. Sin embargo el mismo Newton nos da la respuesta a este problema. Puesto que los datos obtenidos mediante la observación no siempre son completos,



más de alguna vez será necesario interpolar con valores del argumento desigualmente espaciados. Para definir un método de esta naturaleza necesitamos, en primer lugar, generalizar el concepto de diferencias finitas a las denominadas DIFERENCIAS DIVIDIDAS.

#### DIFERENCIAS DIVIDIDAS

Supongamos que se tiene una función definida mediante la tabla:

x	$x_0$	$x_1$	$x_2$	.....	$x_n$
Y	$Y_0$	$Y_1$	$Y_2$	.....	$Y_n$

donde la diferencia entre dos puntos consecutivos cualesquiera es distinto de cero; esto es  $x_{i+1} - x_i > 0$ . Los cocientes:

$$[x_i, x_{i+1}] = \frac{Y_{i+1} - Y_i}{x_{i+1} - x_i}, \quad (i=0, 1, 2, \dots) \text{ se denominan diferencias}$$

divididas de primer orden. Por ejemplo,  $[x_0, x_1] = \frac{Y_1 - Y_0}{x_1 - x_0}$ ,

$$[x_1, x_2] = \frac{Y_2 - Y_1}{x_2 - x_1} \text{ y } [x_2, x_3] = \frac{Y_3 - Y_2}{x_3 - x_2}, \text{ etc, son algunas diferencias}$$

divididas de este orden.

Análogamente se definen las diferencias divididas de segundo orden

$$[x_i, x_{i+1}, x_{i+2}] = \frac{[x_{i+1}, x_{i+2}] - [x_i, x_{i+1}]}{x_{i+2} - x_i}, \quad (i=0, 1, 2, \dots)$$

Por ejemplo  $[x_0, x_1, x_2] = \frac{[x_1, x_2] - [x_0, x_1]}{x_2 - x_0}$  es una diferencia

dividida de segundo orden. Las diferencias divididas de tercer orden se definen en la forma siguiente:

$$[x_i, x_{i+1}, x_{i+2}, x_{i+3}] = \frac{[x_{i+1}, x_{i+2}, x_{i+3}] - [x_i, x_{i+1}, x_{i+2}]}{x_{i+3} - x_i}$$

Por ejemplo  $[x_1, x_2, x_3, x_4] = \frac{[x_2, x_3, x_4] - [x_1, x_2, x_3]}{x_4 - x_1}$  es una

diferencia dividida de tercer orden. En general las diferencias divididas de orden "n" se hallan a partir de las diferencias de orden (n-1) mediante la relación de recurrencia que se ha aplicado. Las diferencias divididas se disponen en una tabla como la siguiente:

TABLA DE DIFERENCIAS DIVIDIDAS

x	y	1a.	2a.	3a.	4a.
$x_0$	$y_0$	$[x_0, x_1]$			
$x_1$	$y_1$		$[x_0, x_1, x_2]$		
$x_2$	$y_2$	$[x_1, x_2]$	$[x_1, x_2, x_3]$	$[x_0, x_1, x_2, x_3]$	
$x_3$	$y_3$	$[x_2, x_3]$		$[x_1, x_2, x_3, x_4]$	$[x_0, x_1, x_2, x_3, x_4]$
$x_4$	$y_4$	$[x_3, x_4]$	$[x_2, x_3, x_4]$		

EJEMPLO. Formar las diferencias divididas de la función definida mediante la tabla siguiente:

x	0	0,2	0,3	0,4	0,7	0,9
y	132.651	148.877	157.464	166.375	195.112	216.000

SOLUCION. Aplicando sucesivamente la relación de recurrencia tenemos:

$$[x_0, x_1] = \frac{148.877 - 132.651}{0,2 - 0} = 81,13$$

$$[x_1, x_2] = \frac{157.464 - 148.877}{0,3 - 0,2} = 85,87$$

$$[x_0, x_1, x_2] = \frac{35,87 - 81,13}{0,3 - 0} = 15,8$$

y así sucesivamente. Los resultados finales de este proceso están especificados en la tabla que sigue:

x	y	1a.	2a.	3a.	4a.
0	132.651				
		81,13			
0,2	148.877		15,8		
		85,87		1	
0,3	157.464		16,2		
		89,11		1	
0,4	166.375		16,7		
		95,79		1	
0,7	195.112		17,3		
		104,44			
0,9	216.000				

Utilizando el concepto de diferencias divididas, podemos representar, ahora, la fórmula de interpolación de Newton para valores del argumento desigualmente espaciados. En efecto:

$$P(x) = y_0 + [x_0, x_1](x-x_0) + [x_0, x_1, x_2](x-x_0)(x-x_1) + \dots + [x_0, x_1, x_2, \dots, x_n](x-x_0)(x-x_1)(x-x_2)\dots(x-x_{n-1})$$

EJEMPLO. La siguiente es una tabla de los sobrevivientes hasta los 15 años de una cohorte de 100.000 nacidos vivos.

x	0	2	5	10
$l_x$	100.000	92.376	90.697	89.870

Utilizando la segunda fórmula de interpolación de Newton, estime los sobrevivientes a la edad de 4 años. ✓

SOLUCION. En primer lugar construiremos la tabla de diferencias divididas.

x	$l_x$	1a.	2a.	3a.	4a.
0	100.000	-3.812			
2	92.376	-559,6	650,5		
5	90.697	-165,4	49,3	-60,12	
10	89.870				

A continuación aplicamos la segunda fórmula de Newton.

$$P(x) = Y_0 + [x_0, x_1](x-x_0) + [x_0, x_1, x_2](x-x_0)(x-x_1) + [x_0, x_1, x_2, x_3](x-x_0)(x-x_1)(x-x_2).$$

Puesto que interpolaremos en  $x=4$  debemos reemplazar,  $x$  por 4, en el polinomio; reemplazar  $x_0, x_1$  y  $x_2$ , por sus respectivos valores; y reemplazar las correspondientes diferencias finitas. En efecto:

$$P(4) = 100.000 + (-3.812)(4-0) + 650,5(4-0)(4-2) + (-60,12)(4-0)(4-2)(4-5), \text{ es decir,}$$

$$P(4) = 100.000 - 15.248 + 5.203,2 + 480,96 = 90.436,16$$

Es decir, se espera que hayan sobrevivido a la edad de 4 años aproximadamente 90.436 personas.

**EJEMPLO.** Un lago es repoblado con 100 peces en el año 1975. La tabla siguiente muestra la observación del crecimiento de esta población a partir de 1980, en los años que se indican.

t	1980	1982	1986	1987
N(t)	1.000	2.000	4.166	4545

Calcular aproximadamente la población en el año 1985. ( $t=0$  en 1980 y  $t=7$  en 1987)

SOLUCION. La tabla de diferencias divididas es la siguiente:

t	N(t)	1a.	2a.	3a.
0	1.000			
2	2.000	500		
6	4.166	514,5	2,41	
7	4.545	379	-27,1	4,21

Aplicando la fórmula de interpolación de diferencias divididas se tiene:

$$P(5) = 1.000 + 500(5-0) + 2,41(5-0)(5-2) + 4,21(5-0)(5-3)(5-6)$$

$$P(5) = 1.000 + 2.500 + 36,15 - 42,1 = 3.494 \text{ peces.}$$

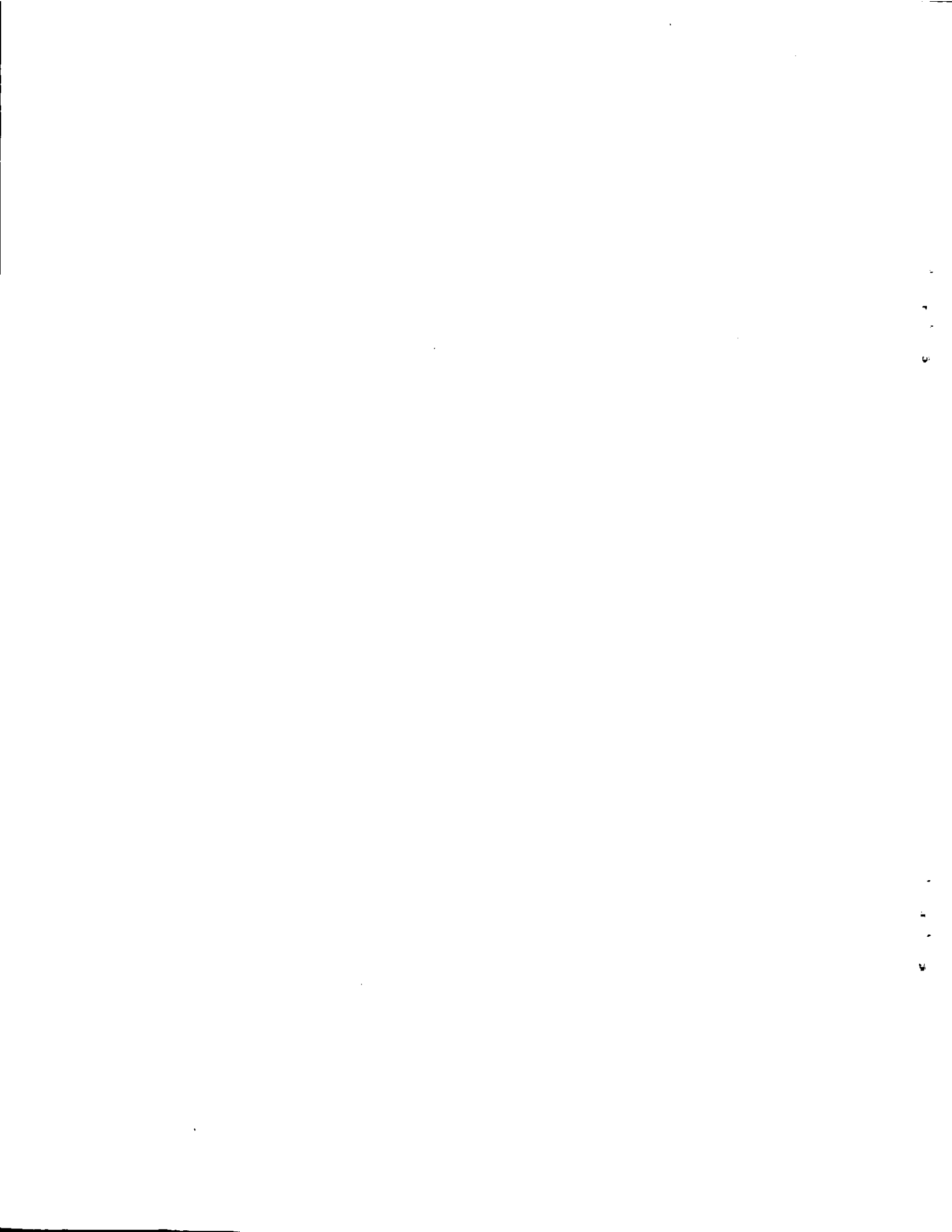
De lo anterior se desprende que en el año 1985 hubo aproximadamente 3.494 peces.

#### INTERPOLACION LINEAL INVERSA

Supongamos que hemos tomado la siguiente información del censo de una ciudad. (x=0 en 1980)

x	0	5	10
y	196.000	210.000	275.000

y deseamos saber en qué año hubo 250.000 habitantes. La respuesta a esta pregunta no es tan inmediata. Notemos que hasta ahora, en general, hemos respondido lo siguiente: ¿Cuál es el valor de la variable dependiente si conocemos el valor de la variable independiente?, ó bien, ¿Cuál es el valor de "y" cuando conocemos el valor de "x"? Ahora debemos dar respuesta a la pregunta, ¿Cuál es el valor de la variable independiente cuando conocemos el valor de la variable dependiente?, dicho de otro modo, ¿Cuál es el valor de x cuando se conoce el valor de y?



Para dar respuesta a esta pregunta utilizaremos, con algunos arreglos, la misma fórmula de interpolación lineal de Newton para valores del argumento igualmente espaciados. Recordemos que la fórmula de interpolación lineal es la siguiente:

$$i) \quad y = y_0 + q\sigma y_0, \quad \text{donde,} \quad ii) \quad q = \frac{x-x_0}{h}$$

Puesto que tenemos que conocer, ahora, el valor de  $x$ , despejamos  $x$  de la ecuación ii). Del despeje resulta que:

$$x = x_0 + qh, \quad \text{donde } q \text{ será ahora, } q = \frac{y-y_0}{\sigma y_0}$$

Estamos en condiciones de dar solución al problema en cuestión. En efecto, construyamos la tabla de diferencias finitas de acuerdo a los datos de la tabla.

x	y	$\sigma y_0$	$\sigma^2 y_0$
0	196.000	14.000	51.000
5	210.000	65.000	
10	275.000		

Considerando los datos dados por el problema y de la tabla de diferencias se tiene:

$$x_0=5, \quad q = \frac{y-y_0}{\sigma y_0} = \frac{250.000 - 210.000}{65.000} = 0,615 \quad \text{y} \quad h = 5$$

Aplicamos la fórmula  $x = x_0 + qh$  resulta que:

$$x = 5 + 0,615(4) = 4 + 2,46 = 6,46 \text{ años.}$$

Esto significa que a mediados de 1987 hubo aproximadamente 250.000 habitantes.

EJEMPLO. La siguiente tabla de valores presenta los sobrevivientes a edades exactas de una generación de 100.000 nacidos vivos entre los 5 y 15 años.

x	5	10	15
$l_x$	92.476	92.321	92.052

¿ A qué edad hubo aproximadamente 92.100 sobrevivientes ?

SOLUCION. La tabla de diferencias finitas es la que sigue:

x	$l_x$	$\sigma y_0$	$\sigma^2 y_0$
5	92.476	-155	-114
10	92.321	-269	
15	92.052		

De los datos del problema y de la tabla de diferencias se tiene la siguiente información:

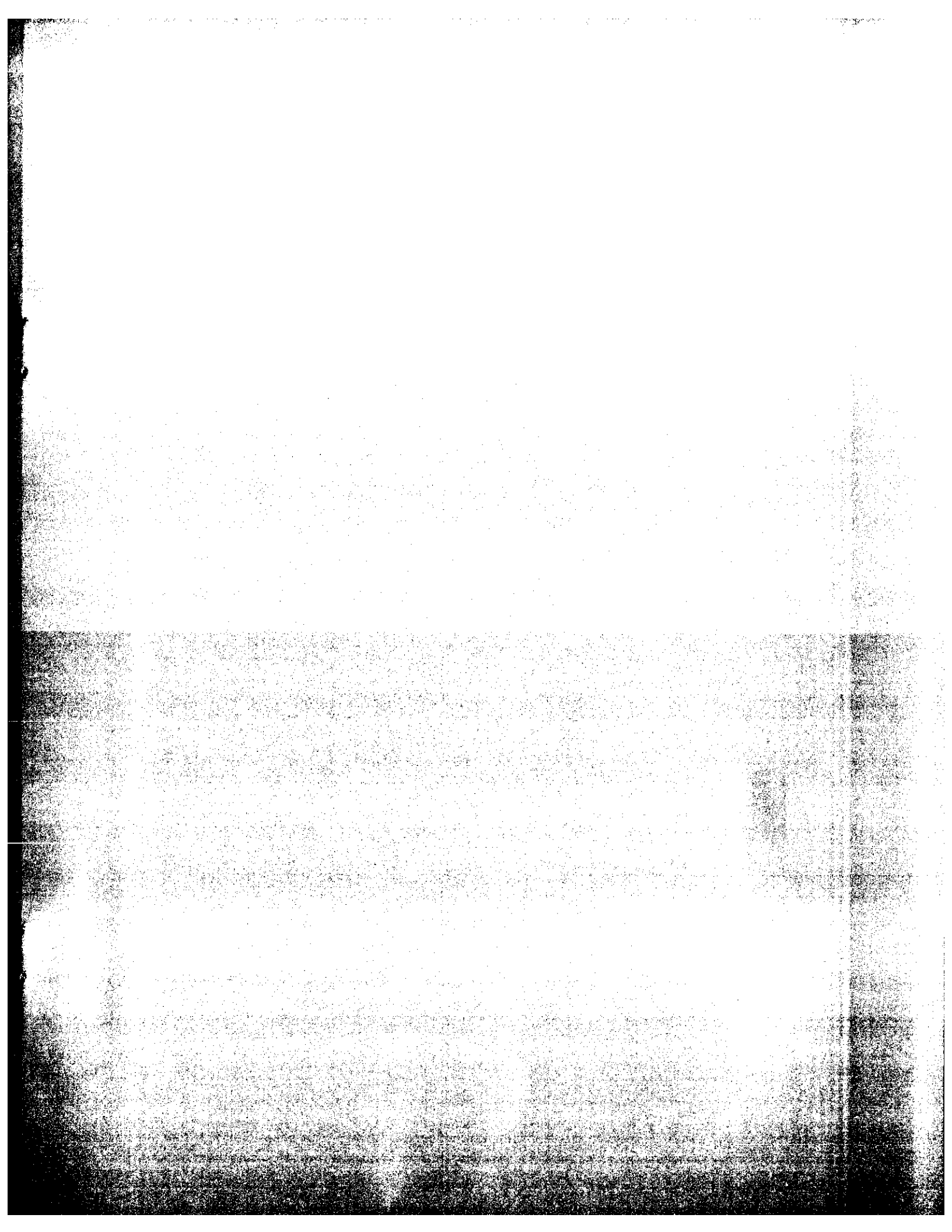
$$x_0 = 10 \quad q = \frac{y - y_0}{\sigma y_0} = \frac{92.100 - 92.321}{-269} = 0,82 \quad \text{y} \quad h = 5$$

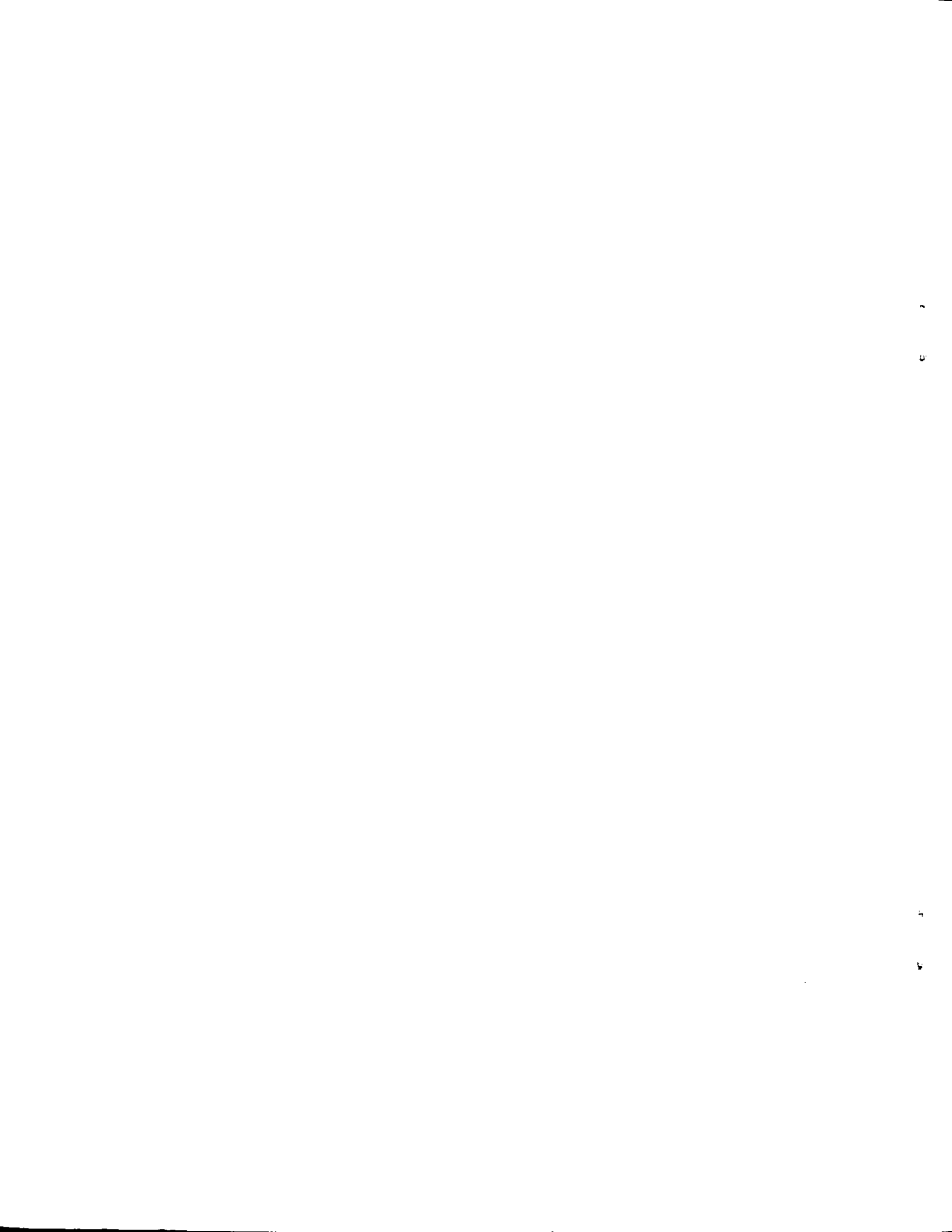
Aplicando la fórmula  $x = x_0 + qh$  resulta:

$$x = 10 + (0,82)(5) = 14,1 \text{ años.}$$

Este resultado indica que hubo 92.100 sobrevivientes a la edad de 14,1 años.







CURSO INTERNACIONAL INTENSIVO DE  
DEMOGRAFIA



ESTADISTICA

Organiza: Facultad de Ciencias Básicas  
Universidad de Antofagasta.

Patrocinan:



UNIVERSIDAD DE ANTOFAGASTA  
FACULTAD DE CIENCIAS BÁSICAS

INSTITUTO NACIONAL DE ESTADÍSTICA



CENRO LATINOAMERICANO DE DEMOGRAFIA



CURSO INTERNACIONAL INTENSIVO DE  
D E M O G R A F I A

APUNTES DE

ESTADÍSTICA

Mg. Jimmy Reyes Rocabado  
Departamento de Matemáticas  
Facultad de Ciencias Básicas  
Universidad de Antofagasta

Mg. Juan Duarte Vargas  
Departamento de Matemáticas  
Facultad de Ciencias Básicas  
Universidad de Antofagasta

## I N D I C E

I.- INTRODUCCION -----	Pag. 1
II.- METODOS DE ENCUESTA-----	Pag. 3
III.- PRESENTACION DE LA INFORMACION -----	Pag. 5
IV.- TASAS, RAZONES Y PROPORCIONES -----	Pag. 21
V.- ESTADISTICA DESCRIPTIVA -----	Pag. 32
VI.- TEORIA ELEMENTAL DE PROBABILIDADES -----	Pag. 45
VII.- DISTRIBUCIONES DE PROBABILIDADES MAS USUALES -----	Pag. 51
VIII.- MUESTREO PROBABILISTICO -----	Pag. 52
IX.- PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA -----	Pag. 59
X.- APLICACIONES DE LA DISTRIBUCION JI-CUADRADO -----	Pag. 77
· APENDICES : TABLAS ESTADISTICAS -----	Pag. 115
REFERENCIAS -----	Pag. 118

## REFERENCIAS

- WAYNE, Daniel " Estadística con aplicaciones a las Ciencias Sociales y a la Educación" Mc. Graw-Hill Colombia 1981.
- D'OTTONE, Horacio " Estadística Nivel Universitario "
- OSTLE, Bernard " Estadística Aplicada " Edit. Limusa Niley S.A. Mexico 1973
- H, Gough " Estadística I " Programa de Postgrado en Programación CELADE-Santiago
- DIVISION DE BIOESTADISTICA Y BIOMATEMATICA " Indicadores de Salud " Escuela de Salud Pública Facultad de Medicina Universidad de Chile

## I. INTRODUCCION

En este curso se estudiará el uso de algunos de los métodos de la estadística en el campo de la demografía y en los estudios sociales.

Empezaremos con una descripción muy breve del campo de estudio que se llama " la estadística ", siguiendo con una caracterización de un estudio estadístico típico de una población grande y de la organización de los datos que provienen de tal estudio. Luego veremos algunos conceptos básicos de la fundamentación teórica de la estadística, empezando con la idea de las probabilidades y del muestreo probabilístico. En los capítulos siguientes veremos varios métodos que permiten el análisis de grandes cantidades de datos en forma resumida, y el cálculo de estimaciones de características específicas de una población. Finalmente veremos los conceptos de distribuciones y de la inferencia estadística en la presencia de errores aleatorios, con determinados métodos de análisis e inferencia muy utilizados en estudios de poblaciones humanas.

¿ Qué es la estadística ?

El " Diccionario de metodología estadística " (G. Gonzalvo, Ediciones Morata, Madrid, 1978, p. 81), define la estadística como sigue :

" Estadística : Técnica o proceso matemático de recogida, descripción, organización, análisis e interpretación de datos numéricos. Como la investigación busca ... una expresión cuantitativa, la estadística constituye un instrumento fundamental de medida y de investigación."

Continúa con dos variedades de estadística :

" Estadística descriptiva (o analítica) : Compendia o simplifica afirmaciones referentes a medidas realizadas sobre una población ... Sirve para la descripción numérica de un grupo particular. Ninguna conclusión va más allá del grupo descrito, y cualquier similitud con los de fuera del grupo no puede ser garantizada. Los datos describen este grupo y solamente éste. "

" Estadística deductiva : Comprende el proceso de muestreo o selección, para su estudio, de un pequeño grupo al que se supone como representativo de otro más numeroso y del cual procede. El pequeño grupo se denomina muestra; el grupo mayor población, conjunto universo."

Notas : 1) en el caso de la estadística deductiva se suele hablar de la inferencia estadística.

2) en este curso se hablará indistintamente de población o universo; sin embargo, el término conjunto se reservará para otro uso.

Campos de aplicación de la estadística. La estadística se puede aplicar a cualquier entidad que se puede describir, aun parcialmente, en términos de características cuantitativas y cuyos elementos pueden ser observados ya sea directa o indirectamente.

Aporte de la estadística al análisis demográfico. Dado que una preocupación muy importante del demógrafo es, en primer lugar, el conocimiento de las características básicas de poblaciones humanas (su número, estructura por sexo y edad, fecundidad y mortalidad, comportamiento migratorio y distribución espacial, actividades económicas y mucho más), las cuales se pueden medir con más o menos precisión, la estadística descriptiva sirve para recolectar, ordenar, resumir e interpretar información sobre estas poblaciones. En segundo lugar, la estadística deductiva (o inferencia) proporciona herramientas para profundizar la interpretación de esa información, y para investigar relaciones entre características, las cuales no siempre se pueden observar directamente. También proporciona indicadores de la calidad o confiabilidad de los datos básicos, además de las medidas resumidas o deductivas que se extraen de ellos.

El siguiente esquema, del libro " Estadística Demográfica " (Universidad de Costa Rica, 1960, Contenido e Introducción ) indica una posible interpretación de la estructura de los estudios demográficos :

" Generalidades sobre recopilación, análisis e investigaciones demográficos

    Recopilación de datos demográficos

        Estadística general de la población

        Estadística de la mano de obra

        Estadísticas vitales

        Estadística administrativa

        Las muestras en la recopilación de datos demográficos

    Análisis de las informaciones demográficas

        Tabulación de los datos

        Análisis estadístico

        Interpretación y presentación

    Investigaciones demográficas básicas "

Métodos estadísticos y métodos científico, puntos de convergencia y de divergencia.

Método científico:

    observación de fenómenos bajo condiciones experimentales controladas



deducción de leyes naturales  
 supuesto que los resultados son reproducibles  
 modelo determinístico

Método estadístico :

observación bajo condiciones parcialmente controladas  
 deducción/inferencia  
 supuesto de una parte determinada y una parte aleatoria ( desconocida)  
 elaboración de modelos para explicar el comportamiento de la parte aleatoria  
 los resultados de un experimento son reproducibles sólo aproximadamente

II. METODOS DE ENCUESTA

En este capítulo queremos considerar las características generales de los estudios que tratan de obtener y analizar información de poblaciones reales, teniendo en cuenta todos los problemas prácticos de estudios de este tipo. Nuestro punto de vista aquí será más la medición a la descripción de la población, que la inferencia en el sentido de la estadística clásica, que veremos en los últimos capítulos de esta asignatura.

Siendo nuestro enfoque principal el campo de los estudios socio-económicos de las poblaciones humanas, empezaremos con una discusión de las etapas de una encuesta típica real. En primer lugar podremos considerar los aspectos generales siguientes :

- El nacimiento de una encuesta
- Las etapas principales de una encuesta
- Las etapas principales de la planificación de una encuesta

El nacimiento de una encuesta

- La necesidad de obtener información
- La especificación de los objetivos
- Consideración de fuentes alternativas de datos
- La necesidad de hacer una encuesta

Definiciones

Encuesta : Una encuesta incluye la recolección de datos sobre una o varias características de algunos elementos de una población con la meta de aprender lo más posible sobre el comportamiento de estas características en la población total.

La encuesta debe desarrollarse teniendo en cuenta los conceptos, métodos y procedimientos claramente definidos. Los datos recolectados

serán recopilados para estimar las características de la población.

Censo : Una encuesta que trata de obtener datos de todos los elementos de la población.

Encuesta por muestro : Una encuesta que busca datos sólo sobre algunos elementos seleccionados de la población.

Normalmente los elementos seleccionados son una fracción muy pequeña de todos los elementos de la población.

Ventajas del muestreo ;

Costos reducidos

Rapidez

Exactitud

Carga reducida para la población

Posibilidad de obtener información detallada

Censo puede ser necesario :

Población pequeña

Necesidad de obtener información para áreas pequeñas

Restricciones legales

Una encuesta puede ser descriptiva o analítica :

Descriptiva : El objetivo primero es de estimar algunas características de la población.

Analítica : El objetivo primero es la verificación (o el rechazo) de una hipótesis, o el análisis de las relaciones entre características.

Las etapas principales de una encuesta

Especificación de los objetivos de la encuesta

Identificación de la población bajo estudio

Preparación del marco muestral

Confección del cuestionario

Diseño muestral

Recolección de datos

Procesamiento de los datos

Control de errores

Estimación

Evaluación y análisis de los datos

Publicación

## Las etapas principales de la planificación de una encuesta

Objetivos de la encuesta  
 Estudio de factibilidad  
 Equipo de proyecto  
 Presupuesto  
 Calendario  
 Definición de la población bajo estudio  
 Datos que se deben recolectar y análisis previsto  
 Precisión exigida  
 Métodos de recolección de datos  
 Trabajo de terreno  
 Métodos de muestreo  
 Métodos de digitación y procesamiento de los datos  
 Métodos de verificación e imputación  
 Métodos de estimación  
 Tablas que se deben producir  
 Control de la no-respuesta  
 Encuesta piloto  
 Ciclo de retroalimentación

### Objetivos de la encuesta

Se definen en término de los datos requeridos para obtener las conclusiones deseadas  
 Importancia de la precisión que se trata de lograr  
 Consecuencias de decisiones malas basadas en datos imprecisos  
 Efectos sobre el presupuesto y sobre el calendario de actividades  
 En realidad, los objetivos se modifican durante la preparación de la encuesta

### Estudio de factibilidad

Normalmente se debería efectuar un pequeño estudio preliminar para poder decidir si la encuesta realmente puede proporcionar la información requerida

### Equipo de proyecto

Es el grupo de personas encargadas de llevar a cabo la encuesta  
 El gerente de proyecto es miembro del equipo  
 Normalmente, el equipo incluye representantes de varios campos especializados:

El campo sustantivo de estudio de la encuesta ( demógrafo, sociólogo, economista, experto en salud pública, ... )

Metodología de encuesta/muestreo (estadístico)

Computación (analista/programador)

Operaciones regionales

Cada persona tiene sus tareas especializadas

Todo el equipo debe mantenerse al día en cuanto a todas las etapas del proyecto

Todos participan en la planificación de cada etapa

#### Presupuesto

Identificación de los recursos necesarios y los recursos disponibles :

Dinero

Personal capacitado

Computadoras y otros equipos

Otro

#### Calendario

Hay que establecer fechas límite para cada etapa mayor del proyecto

El gerente de proyecto controla estas fechas para respetar los límites globales de tiempo para terminar el proyecto

#### Definición de la población bajo estudio

Es necesario definir con cuidado la población bajo estudio, y la población para la cual se quiere tomar decisiones basadas en los resultados de la encuesta

#### Datos que se deben recolectar y análisis previsto

Hay que decidir exactamente qué datos se buscan

¿ Cómo se van a utilizar ?

¿ Cuáles resultados se publicarán ?

¿ Cuáles resultados serán sólo para análisis interno ?

Importancia de una boleta bien preparada

#### Precisión exigida

Es necesario decidir qué nivel de precisión es suficiente para el análisis y para las decisiones sobre políticas

Se habla de la confiabilidad de los resultados

El nivel de la confiabilidad está estrechamente ligado al tamaño de la muestra, entonces a los costos del proyecto

#### Métodos de recolección de datos

Generalmente, es posible elegir uno entre varios métodos disponibles para la recolección de los datos :

Entrevista personal

Entrevista por teléfono

Encuesta por correo

Medición directa

Archivos administrativos

Combinaciones de estos métodos

Trabajo de terreno

Todas las operaciones de terreno deben ser planificadas con gran cuidado y bien controladas sobre la marcha para garantizar la obtención de datos completos y correctos

Capacitación del personal :

- Entrevistadores
- Supervisores
- Personal que digite los datos

Verificación de los datos sobre la marcha y control de las respuestas

Verificación que la última etapa del muestro se hizo correctamente  
Visitas repetidas para minimizar la no-respuesta

Métodos de muestreo

Hay varios métodos que se pueden utilizar para seleccionar la muestra  
Se trata de elegir lo más apropiado, tomando en cuenta :

- Consideraciones operacionales
- Costos
- Información necesaria para la preparación del marco muestral

Entre los métodos que parecen factibles, se compara el tamaño de la muestra que requiere cada método para lograr la precisión requerida con el costo de cada alternativa

Métodos de digitación y procesamiento de los datos

Hay varios métodos para captar los datos :

- Codificación manual
- Tarjetas perforadas
- Digitación y verificación ("key-edit")
- Lectura óptica

Hay varios métodos de procesamiento de los datos :

- Tabulación manual
- Procesamiento por computadora :
  - con programas escritos para el proyecto específico
  - con "paquetes software" para análisis estadístico (SPSS,SAS MINITAB, SL-MICRO, EMDP, PANDEM, CHECKEDIT,...)

Los métodos elegidos para estas dos etapas deben ser compatibles

Métodos de verificación e imputación

Verificación (control) :

- Métodos que tratan de identificar los datos incompletos, incorrectos o incoherentes

#### Imputación :

Métodos para corregir los problemas identificados durante la etapa de verificación

En ambos casos hay una selección de métodos

Es necesario especificar completamente los pasos operacionales de la implementación del método elegido para los datos de esta encuesta

Normalmente es necesario prever la preparación de programas para la computadora

#### Métodos de estimación

Son procedimientos matemáticos para calcular las estimaciones de las características de la población a partir de los datos muestrales

Incluyen procedimientos para la estimación de la confiabilidad de esas estimaciones

La selección del método es normalmente dictada por la selección del plan de muestreo

#### Tablas que se deben producir

Tablas que se publicarán como resultados principales de la encuesta

Tablas que se prepararán sólo para ciertos usuarios

Datos que se conservarán para ser analizados por el equipo de proyectos o sus colegas

#### Control de la no-respuesta

Hay varios métodos para reducir la tasa de no-respuesta :

Carta de presentación

Propaganda

Visitas repetidas

En última instancia, es posible ajustar los factores de ponderación para corregir el efecto de la no-respuesta sobre las estimaciones

#### Encuesta piloto

"Pre-test" de algunos de los pasos de la encuesta, sobre todo la boleta

o

Encuesta piloto de todas las etapas

El pre-test puede ser solamente cualitativo, por ejemplo tratando de identificar problemas con un borrador de la boleta, o sirviendo para la formación de los entrevistadores

Una encuesta piloto debe normalmente comprobar todos los procedimientos de la encuesta, hasta la producción de las tablas

A menudo la encuesta piloto proporciona resultados preliminares cuantitativos que permiten mejorar la alocación de los recursos para la encuesta principal

Si este tipo de encuesta se hace por primera vez, la encuesta piloto es imprescindible

Estos dos tipos de prueba pueden permitir mejorar de manera significativa la encuesta principal

Se obtiene información útil sobre los costos y el tiempo que realmente hay que prever para las etapas mayores

#### Ciclo de retroalimentación

Todo lo que se aprende durante la planificación y las operaciones de la encuesta, proporciona información importante que permite mejorar los métodos cuando la encuesta se repita, o cuando se haga otra encuesta del mismo tipo general

La redacción de documentos claros explicando los métodos utilizados y las lecciones aprendidas, es esencial para que esta información sea útil más tarde

### III. PRESENTACION DE LA INFORMACION

#### 1.- NIVELES DE MEDICION

Los niveles de medición son diferentes escalas que se usan para medir o describir objetos, individuos u otras entidades y se dividen en :

	nominal	
Escalas	ordinal	Discreta
	De Intervalos	Continua

Escala Nominal : Es aquella que mide o describe asignando " nombres " a los objetos, individuos u otras entidades y se usan cuando se quiere medir una cualidad o atributo.

Por ejemplo : Profesión ( Médico, Ingeniero, Profesor, etc. ) Color de pelo ( rubio, negro, etc. ). Estado Civil ( casado, soltero, viudo ).

Escala Ordinal : Es aquella que jerarquiza u ordena permitiendo indicar la posición relativa de los distintos elementos clasificados.

Por ejemplo : Rendimiento de un curso  
Grado de dificultad de una prueba  
El desarrollo de un país

#### Escalas de Intervalos

a) Discreta : es aquella que puede tomar sólo ciertos valores en el

intervalo considerado y no admite valores intermedios. Por ejemplo número de hijos, número de estudiantes, etc.

- b) Continua : es aquella que puede tomar cualquier valor en un intervalo dado. por ejemplo : peso, estatura, rendimiento, etc.

## 2.- TABLAS ESTADISTICAS

Las tablas estadísticas sirven para presentar los datos numéricos obtenidos en algún estudio en forma clara y ordenada.

Las etapas principales en la construcción de una tabla son:

- a) Definir los propósitos de la tabla : según los propósitos se distinguen dos tipos de tablas :
- i) Tablas de frecuencias, donde el material se clasifica según un solo criterio.
  - ii) Tablas de asociación, donde se desea mostrar la relación entre dos o más variables en las unidades de observación.

La definición de los propósitos ayuda a determinar los criterios de clasificación a usar en la tabla y el sentido en que deben analizarse los datos.

Por ejemplo, i) Si el propósito es mostrar las edades de un grupo de estudiantes, sólo se empleará un criterio de clasificación la edad y se construirá una tabla de distribución de frecuencias.

ii) Si el propósito es en cambio, mostrar la relación que existe entre la edad y el rendimiento del estudiante, entonces se hará una tabla de asociación con dos criterios de clasificación, la edad y el rendimiento.

Observación, Una manera práctica de definir los propósitos de la tabla es a través de la formulación de la o las preguntas que se intenta contestar con la tabla.

Por ejemplo, para el ejemplo i) anterior se podría preguntar ¿Cuál es la distribución por edad de los estudiantes? y en el caso ii) ¿Hay relación entre la edad de los estudiantes y su rendimiento?

- b) Colocar un Título

Las tablas deben tener un título que especifique :

- i) Qué se presenta :

Por ejemplo : estudiante, profesores, rendimiento, etc.

- ii) Cómo se clasifican las unidades de observación :

Por ejemplo, estudiantes según edad, pruebas, según resultado estudiantes según rendimiento, etc.



- iii) Donde fueron obtenidos los datos :  
 Por ejemplo : Departamento de Matemáticas del Liceo  
 Liceo de Hombres de Antofagasta, etc.
- iv) Cuando se registraron los datos
- c) Asignar las escalas de clasificación a filas y columnas
- i) Cuando hay un solo criterio de clasificación de las observaciones, se colocará la escala de clasificación en la primera columna.

Ejemplo clasificación por edad

EDADES	Nº OBSERVACIONES
0 - 4	10
5 - 9	14
10 - 14	25
TOTAL	49

- ii) Cuando hay mas de un criterio se preferirá colocar la escala con mayor número de grupos o categorías en la primera columna

Ejemplo Al clasificar por edad y rendimiento a un grupo de estudiantes

EDAD	RENDIMIENTO			TOTAL
	BUENO	REGULAR	MALO	
0 - 4	13	20	23	66
5 - 9	11	19	29	58
10 - 14	14	23	37	74
TOTAL	38	61	99	198

- iii) Cuando hay dos escalas de clasificación y una se refiere a los antecedentes y otra a las consecuencias, es preferible colocar los antecedentes en la primera columna y las consecuencias en la fila superior.

Ejemplo Para estudiar la relación entre tipo de enseñanza y rendimiento en una prueba.

TIPO DE ENSEÑANZA	RENDIMIENTO			
	BUENO	REGULAR	MALO	TOTAL
TRADICIONAL				
PERSONALISADA				
TOTAL				

Ejemplo resuelto :

Construir una tabla para mostrar si la postulación a la Universidad depende del tipo de educación y del sexo.

TITULO : Postulación a las Universidades según tipo de educación y sexo.

Servicio de Selección y Registro de Estudiantes Universidad de Chile 1985.

Escala de clasificación

Son 3 : postulación, tipo de educación y sexo.

El tipo de educación se ubica en la primera columna, pues es posible dividirlo en un número mayor de grupos, los otros dos criterios se colocarán en las filas superiores.

Colocación de datos numéricos

Para cada sexo y para cada tipo de educación el desenlace es postula o no postula; como el interés de nuestro estudio es la postulación sólo se colocará el N° de postulante.

TIPO DE EDUCACION	FEMENINO		MASCULINO			
	POSTULAN	TOTAL	POSTULAN	TOTAL	TOTAL	
Cient. Hum. Noct.	384	65,1	590	1.010	69,9	1.445
Comercial	1.006	71,5	1.407	916	75,0	1.221
Industrial	145	78,4	185	3.924	74,3	5.281
Técnica	452	75,2	601	40	85,1	47
Agrícola	11	100,0	11	34	72,3	47
Cient. Hum. Diur.	14.699	80,9	30.514	28.170	81,9	34.378
TOTALES	26.697	80,2	33.338	34.094	80,4	42.419

## Ejercicios

- 1.- Indique en los datos que siguen , qué escala de clasificación les corresponde ( marque con una x )

### ESCALA DE CLASIFICACION

DATOS	NOMINAL	ORDINAL	DISCRETA	CONTINUA
Tamaño grupo familiar				
Previsión social				
Número de hijos vivos				
Nivel socio-económico				
Actividad económica				
Ingresos económicos				
Nivel educacional				
Nº de piezas en la vivienda				
Material de construcción				
Sexo				
Talla				
Desarrollo psicomotor				
Diagnóstico de la enfermedad				
Temperatura				
Presión arterial				
Gravedad de la enfermedad				
Nº de días de hospitalización				

- 2.- Construya una tabla estadística para mostrar la distribución de frecuencias de diagnóstico en estudiantes con problemas de aprendizaje
- 3.- Construya una tabla para mostrar el efecto de dos tratamientos en la evolución del aprendizaje en estudiantes con retardo mental.  
( El tratamiento puede ser con estímulos y sin estímulos y la evolución del aprendizaje se mide si el estudiante mejora, sigue igual o empeora )
- 3.- Construcción de una tabla de distribución de frecuencias para el caso de una escala de intervalos continua

#### Reglas generales :

- i) Se ordenan los datos en forma creciente
- ii) Determinar el mayor y el menor entre los datos recopilados y así

vi) Colocación de un título, el cual debe contener en forma clara y lo más completa posible la información necesaria para su fácil comprensión

TIPOS DE GRAFICOS

a) Gráficos lineales

Se utilizan para presentar información, especialmente en el caso de escala de intervalos continuo. Consiste en un par de ejes cartesianos donde en la abscisa se ubican los valores de la variable y en la ordenada los cambios de la misma mediante una escala de tipo aritmético. Cada pareja de datos se representa por un punto. Luego habrán tantos puntos como parejas de datos, los que se unen mediante segmentos rectilíneos.

Ejemplo Tasa de mortalidad infantil. Santiago 1940 - 1967



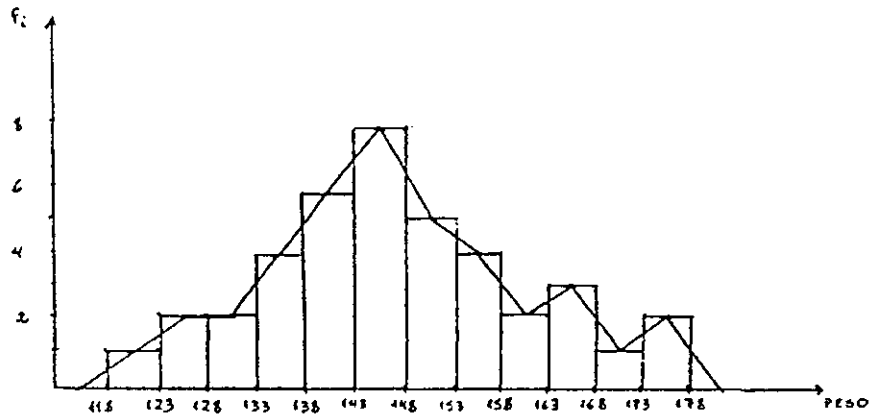
b) Histogramas

Se utilizan de preferencia cuando en el eje horizontal aparecen una escala de intervalos continua. Está formado por rectángulos de igual base y cuya altura es proporcional a la frecuencias respectiva conforme a la indicación en el eje vertical.

Para su mejor presentación es conveniente que todos los intervalos de clase sean de igual tamaño.

Algunas veces se acostumbra a unir los puntos medios de la parte superior de cada rectángulo mediante segmentos rectilíneos; el gráfico así formado se denomina polígono de frecuencia

Ejemplo Con respecto al ejemplo del peso de 40 estudiantes



los histogramas o gráficos de Barras pueden adoptar varias formas  
Así tenemos :

i) Gráfico de barra simples

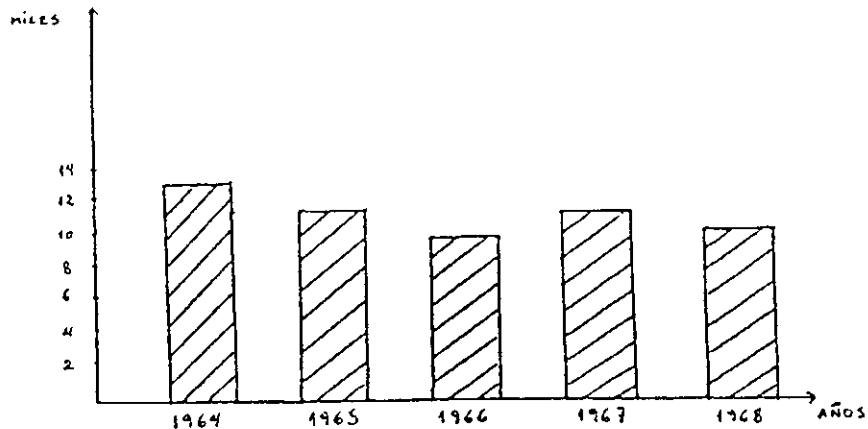
Se utilizan para comprobar cifras totales expresadas en unidades originales.

Cada barra tiene una longitud proporcional a la frecuencia de la variable, la que puede leerse en la escala vertical.

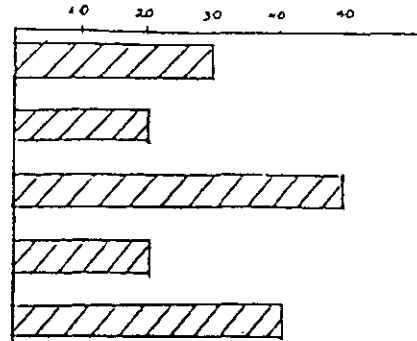
Cuando las cifras se refieren a algún atributo, las barras pueden dibujarse horizontalmente con el objeto de colocar fácilmente a la izquierda de la figura la leyenda correspondiente .

Ejemplos

1.- Número de estudiantes atendidos en colegio x por año



2.- Distribución de estudiantes según tipo de educación

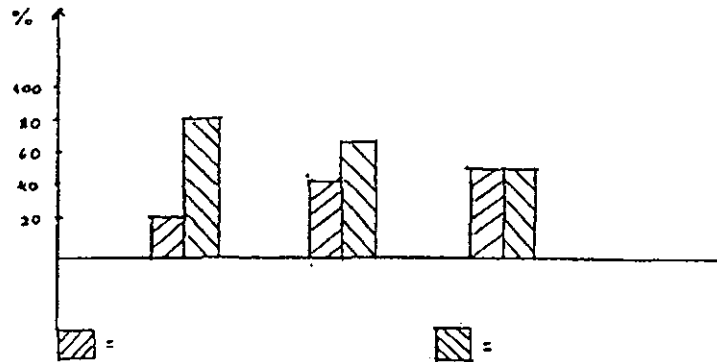


ii) Gráfico de Barras Agrupados

Se utilizan generalmente para comparar dos variables según la frecuencia con que se presentan ( frecuencia porcentual )

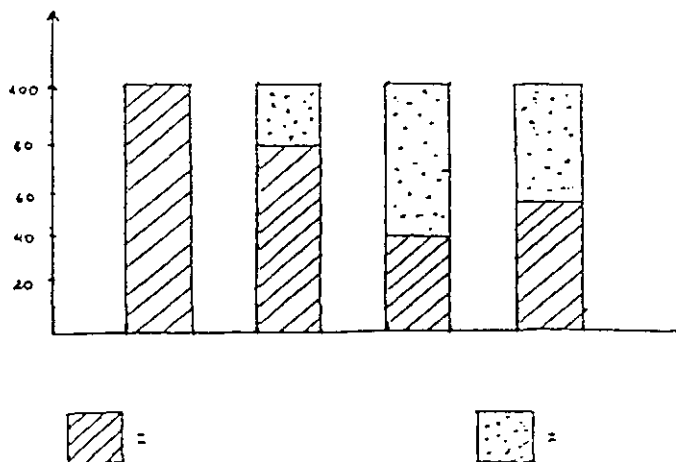
Por Ejemplo :

Porcentaje de estudiantes accidentados según gravedad de las lesiones en diferentes edades.



iii) Gráficos de Barras Subdivididos

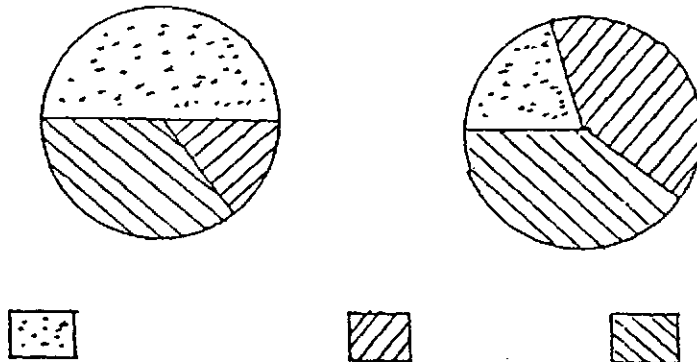
Es un gráfico muy apropiado para mostrar la composición proporcional de distintas categorías. No es conveniente hacer mas de tres subdivisiones.



c) Gráficos Sectoriales

Se utilizan para los mismos casos que los gráficos de barras, con la limitante que toda frecuencia debe expresarse como proporción del total. Esta proporción determina el  $\%$  con respecto a los  $360^\circ$  de la circunferencia total que debe limitar el sector que representa la frecuencia correspondiente.

Ejemplo Resultados obtenidos con tratamientos A y B



Ejercicios Propuestos

1.- Construya el (los) gráfico (s) adecuado para presentar la información que aparece en las tablas siguientes y haga una pequeña interpretación de ellos.

i) Retiro de estudiantes del Ciclo Básico según causa Area Occidente 1972.

CAUSAS	N°	%
SALUD	93	12,0
ECONOMICOS	236	30,4
TRASLADO	57	7,3
OTRAS	391	50,3
TOTAL	777	

ii) Rendimiento académico por Departamento del Liceo X 1980

DEPARTAMENTO	RENDIMIENTO	
	HORAS HABILES	HORAS TRABAJADAS
MATEMATICAS	3.7	4.1
FISICA	6.9	7.1
BIOLOGIA	5.3	5.5
QUIMICA	4.8	5.1
ARTES	6.6	6.7
ED. FISICA	2.6	7.8
MUSICA	3.3	3

2.- Los datos que aparecen a continuación corresponden al N° de personas del grupo familiar de algunas familias .

Construya una tabla de frecuencia y clasifique estos datos en ella y anexe el gráfico correspondiente.

6	7	8	4	6	7	7	5	6	4
4	8	11	3	6	6	7	4	3	5
5	4	2	3	3	3	4	5	8	6
4	8	5	5	5	4	2	5	3	4
4	8	6	7	5	5	10	8	9	3
5	5	5	5	2	11	8	8	4	3
4	7	2	4	6	9	4	5	6	2
3	3	7	7	4	5	8	3	7	6
7	6	4	7	3	9	4	3	7	3
3	10	6	6	10	6	12	8	4	4



#### IV TASAS, RAZONES Y PROPORCIONES

##### Cifras absolutas y frecuencias relativas

Las estadísticas que resultan de las tabulaciones de diferentes tipos de datos (nacimientos, defunciones, casos de enfermedad, consultas, egresos hospitalarios, etc.) proporcionan números absolutos que son muchas veces utilizables directamente en Salud Pública. Por ejemplo, el número de consultas otorgadas en un consultorio externo permite al administrador en salud estimar la cantidad de recursos necesarios para dar una atención suficiente; el número de nacimientos es un dato valioso para los programas de atención materno-infantil; el número de egresos de un hospital muestra el volumen de hospitalización y sirve para calcular costos y rendimientos.

Sin embargo, a pesar de la importancia de las cifras absolutas, son las frecuencias relativas las que tienen una mayor utilidad. Bajo esta denominación se incluyen las tasas, proporciones, porcentajes y simples razones. Las frecuencias relativas tienen la ventaja de facilitar la presentación de las relaciones que existen entre dos o más datos y hacer más sencilla la comparación de resultados.

##### 1. Razones

Son cuocientes entre dos cantidades de igual o distinta naturaleza. Indican cuantas veces sucede el hecho que está en el numerador con respecto al hecho que está en el denominador.

Ejemplo : Razón de masculinidad =  $\frac{\text{Nº de hombres}}{\text{Nº de mujeres}}$

Indica cuántos hombres hay por cada mujer. Si se amplifica por 100, se sabrá cuantos hombres hay por cada 100 mujeres, en Chile 1982 había 96 hombres por cada 100 mujeres.

$$\text{Chile 1982} = \frac{5.521.067}{5.754.373} \times 100 = 95.9$$

##### Otro ejemplo :

En el programa de atención maternal se desea comparar la relación entre controles y consultas de morbilidad otorgadas en dos Servicios de Salud en 1982.

Servicio de Salud	Atención maternal	
	Controles	Consultas morbilidad
Oriente	72.154	72.568
Sur	72.029	87.041

El examen de estas cifras absolutas hace un poco difícil la comparación. En una forma gruesa se puede decir que ambos Servicios dieron igual número de controles y que, en cambio, el número de consultas por morbilidad fue muy superior en el Servicio Sur. Resulta más clara la comparación si se calculan los cuocientes entre el número de controles y el número de consultas en cada uno de los Servicios.

Servicio Oriente  $72.154/72.568 = 1$  control por cada consulta.

Servicio Sur  $72.029/87.041 = 0.8$  controles por cada consulta.

Se establece que el Servicio Oriente ha dado más controles por consulta que el Servicio Sur.

## 2. Proporciones

Son cuocientes entre dos cantidades de igual naturaleza. Describen la fracción que una serie de sucesos que figuran en el numerador representa con respecto al total de sucesos de igual índole. Cuando el resultado de este cuociente se multiplica por 100 resulta un porcentaje, que es la forma habitual de calcular esta frecuencia relativa.

Ejemplo: En Chile en 1982 el Sistema Nacional de Servicios de Salud controló el estado nutricional de 1.160.813 niños menores de 6 años. En el mismo año la Región Metropolitana controló 390.464 niños de igual edad. Como la Región Metropolitana es una parte del Sistema Nacional se puede calcular el porcentaje que representan los controles de esta Región con respecto al total del país:

$$\frac{390.464}{1.160.813} \times 100$$

De este modo se sabe que 34% del total de niños menores de seis años un control nutricional en el país, pertenecen a la Región Metropolitana.

Es importante insistir que tanto los hechos que figuran en el numerador como los del denominador deben ser de igual naturaleza. De este modo el resultado expresa la importancia relativa que el dato del numerador tiene con respecto al total.

Los porcentajes tienen la ventaja de permitir una comparación fácil de series que tienen totales diferentes, al referirlos a una base común que en este caso es 100. Si suponemos dos Provincias en que se desea conocer si la mortalidad del menor de 28 días es diferente en importancia con respecto al total de niños menores de 1 año, es más sencillo calcular los porcentajes que representan las defunciones de menores de 28 días con respecto al total de defunciones de menores de 1 año.

Provincia	Defunciones de menores de 1 año	Defunciones de menores de 28 días	
		Nº	%
Concepción	502	242	48,2
Bio-Bio	287	131	45,6

En la provincia de Concepción las defunciones de menores de 28 días representan el 48,2% del total de defunciones infantiles, en cambio en la provincia de Bio-Bio representan el 45,6%.

Limitaciones de los porcentajes y necesidad del cálculo de tasas

A pesar de su utilidad, los porcentajes tienen limitaciones. Si se estudian, por ejemplo, los muertos por accidentes en dos grupos de edades en un país X nos encontramos con lo siguiente:

Grupo de edad	Total de defunciones	Defunciones por accidente	
		Nº	%
15 a 24	48.999	12.763	26,0
65 a 74	306.025	11.425	3,7

*P. V. M.*

En este caso podría concluirse que los accidentes son un peligro más serio para los jóvenes, en los que más de una cuarta parte de las defunciones se debe a accidentes, que para las personas de mayor edad, en las que los accidentes causan menos del 4% de las defunciones.

Las cifras anteriores no expresan realmente el riesgo de morir por accidente, sino la importancia relativa que esta causa tiene en el total de defunciones de cada grupo de edad. El conocimiento del riesgo no se obtiene con el cálculo de los porcentajes; para ello hay que introducir en la comparación un elemento importante que es la población expuesta al riesgo de sufrir accidentes. El resultado que se obtiene al dividir el número de muertes debidas a accidentes por la población expuesta al riesgo de sufrir un accidente es lo que se denomina tasa de mortalidad por accidente.

### 3. Tasas

Una tasa es un cociente formado por tres elementos :

- Un numerador, que consiste en el número de veces que ocurrió un determinado hecho en un período de tiempo dado y en un área determinada. Por ejemplo, el número de casos de una enfermedad que se registró en un área durante un año.
- Un denominador, que es la población expuesta al riesgo de que le suceda el fenómeno que aparece en el numerador.
- Una constante por la cual se multiplica el cociente. Debido a que el cociente resultante en una tasa es siempre de valor inferior a la unidad, éste se multiplica por 100, 1.000, 10.000 ó 100.000 de modo de tener cifras superiores a la unidad, lo que facilita la interpretación. En efecto, es más fácil entender que la tasa de mortalidad de una región es 8 por 1.000 habitantes que decir que es 0,008 por habitante.

#### Requisitos generales de las tasas

Es necesario que en una tasa haya concordancia entre el numerador y el denominador en tres aspectos importantes: la naturaleza del hecho, la zona geográfica y el período de tiempo dentro del cual ocurre el hecho.

En relación con la naturaleza del hecho, debe usarse en el denominador la población de la cual haya emanado el hecho del numerador. Así, no podríamos tener una tasa de mortalidad por cáncer de la próstata si en el denominador figura la población femenina.

El área geográfica debe ser la misma para el numerador que para el denominador.

Con respecto al tiempo, las tasas se calculan generalmente sobre una base anual. Se presenta un problema en cuanto al denominador de la tasa, ya que debido a que la población varía a lo largo del año, pueden hacerse distintas estimaciones de ella. Si la población se estima al comienzo del período no representa toda la población expuesta ya que en esta población no figuran por ejemplo, los niños que nacerán durante el año. Si la población que se usa es la estimada al final del año sucede lo contrario, ya que no aparecerán en ella los que han fallecido y los que han emigrado en el curso del año.

De aquí que es de uso habitual como representativa de la población media expuesta al riesgo la estimación a mitad del período, es decir, al 30 de junio del año en estudio.

Tipos de tasas

En general pueden distinguirse dos tipos principales de tasas:

- a) tasas crudas o brutas.
- b) tasas específicas.

Cuando en el denominador figura el total de la población se habla de tasas crudas porque no se consideran características como edad, sexo, etc. Es una medición gruesa de la fuerza de ocurrencia de un hecho.

Cuando en el denominador se usa sólo cierto sector de la población, por ejemplo, la población de 20 a 25 años (en el numerador debe figurar el hecho referido que afecta sólo a este grupo de edad) se habla de tasas específicas. Estas tasas son más refinadas y miden con mayor exactitud el riesgo que se desea conocer, ya que en general los riesgos son diferentes según las características de las personas.

Por ejemplo, la mortalidad es muy diferente en algunos grupos de edad y la tasa cruda es sólo una especie de promedio de las diferentes tasas específicas. A veces se habla impropriamente de que una tasa es específica. Tal es el caso de la tasa de mortalidad por una causa determinada, por ejemplo tuberculosis. Si en el numerador figuran todas las defunciones por tuberculosis en el denominador debe estar toda la población y es por lo tanto una tasa cruda por una causa específica.

Las tasas que habitualmente se usan en Salud Pública se refieren a la mortalidad, la morbilidad, la letalidad y la fecundidad.

3.1. Tasa bruta de mortalidad

Su numerador incluye la totalidad de las defunciones de ambos sexos, de todas las edades y por todas las causas, registradas a lo largo de un año calendario en un área determinada. Su denominador es la población total de esa misma área estimada a mitad de período, es decir, al 30 de junio del mismo año. Tal como ocurre con todas las tasas de mortalidad, debido a que en la población expuesta al riesgo de morir sólo algunos individuos han muerto en el término del año calendario, el denominador es siempre mayor que el numerador y, para obtener cifras enteras es necesario amplificar el cociente entre defunciones y población por una constante que, en el caso de la tasa bruta es 1.000.

Tasa bruta de mortalidad =

$$= \frac{\text{No total de defunciones en un área y año determinados}}{\text{Población total del área al 30 de junio de ese año}} \times 1.000$$

Según causa

Tasa de mortalidad por causa =

$$= \frac{\text{Defunciones por una causa en un área y año determinados}}{\text{Población total al 30/junio de ese año y área}} \times 100.000$$

El denominador de las tasas por causa, en general, es la población total y por consiguiente se trata de tasas crudas por una causa o grupo de causas específicas.

La construcción de estas tasas implica separar el conjunto de todas las muertes diversos sub conjuntos atendiendo a la causa de muerte. Dichas muertes, si no hay otra especificación adicional, incluyen las defunciones de cualquier edad y ambos sexos que han ocurrido por una misma causa o grupo de causas.

Debido a la necesidad de disponer de tasas por causas de muerte cuya magnitud en la población puede ser muy pequeña y a fin de que la magnitud de las tasas de mortalidad por las diferentes causas sea fácilmente comparable, la constante que en ellas se utiliza es 100.000.

Tasa de mortalidad materna =

$$= \frac{\text{Muertes debidas a complicaciones del embarazo, parto o puerperio}}{\text{Nacidos vivos en ese año y área}} \times \frac{1.000}{10.000}$$

Se denominan muertes maternas aquellas cuya causa está relacionada con complicaciones del embarazo, parto o puerperio y ellas constituyen el numerador de la tasa.

Su denominador podrían ser las mujeres entre 15 y 49 años pero el riesgo específico que indica el numerador sólo afecta a aquellas que en dicho año han tenido un embarazo, por lo tanto lo más adecuado sería colocar el número de embarazos. Como habitualmente no se dispone de información fidedigna respecto a este dato, se ha convenido internacionalmente utilizar como denominador el número de nacidos vivos del mismo año en que sucedieron las muertes del numerador.

La tasa de mortalidad materna se define como la relación entre el número de defunciones por causas relacionadas con las complicaciones del embarazo, parto o puerperio ocurridas en un año y área dadas y el número de nacidos vivos en el mismo año y área. Se puede expresar por 1.000 ó por 10.000.

3.2. Tasas específicas de mortalidad

Según sexo:

El riesgo de morir difiere según el sexo. Por ello es conveniente medir por separado la mortalidad de hombres y de mujeres.

Tasa Mortalidad Masculina =

$$= \frac{\text{Defunciones masculinas en un área y año determinados}}{\text{Población masculina al 30/VI de ese año y área}} \times 1.000$$

Tasa Mortalidad Femenina =

$$= \frac{\text{Defunciones femeninas en un área y año determinados}}{\text{Población femenina al 30/VI de ese año y área}} \times 1.000$$

Igual que la tasa bruta de mortalidad, ambas tasas se amplifican por 1.000.

Debido a que sus denominadores son diferentes estas dos tasas no se pueden sumar directamente para reconstruir la tasa bruta de mortalidad.

Según edad :

La mortalidad difiere marcadamente según la edad. Por eso corrientemente la medición de la mortalidad requiere medir el riesgo de muerte por edades. Al elaborar las tasas de mortalidad por edad puede llegarse a tal grado de especificación que los subconjuntos de defunciones incluyan sólo edades simples, es decir, se elabore una tasa para cada año de edad. Sin embargo, lo habitual es que se trabaje con grupos de edades, usándose frecuentemente grupos quinquenales de edad o bien grupos de mayor amplitud. Sólo para las edades más jóvenes, en que el riesgo de morir cambia más rápidamente con la edad, está justificado construir tasas de mortalidad por edades simples o aún por intervalos que sean menos amplios que 1 año.

Tasa de mortalidad por edad =

$$= \frac{\text{Defunciones de un grupo de edad en un área y año determinados}}{\text{Población de ese grupo de edad al 30/VI de ese año y área}} \times 1.000$$

Todas las tasas de mortalidad por edad se amplifican por 1.000. Estas tasas se pueden calcular separadamente para cada sexo. En tales casos la doble especificación de sexo y edad debe hacerse tanto para las defunciones como para la población. Ejemplos:

Tasa mortalidad masculina de 20 - 24 años =

$$= \frac{\text{Defunciones masculinas de 20-24 en un área y año determinados}}{\text{Población masculina de 20-24 años al 30/VI para ese año y área}} \times 1.000$$

Un caso especial dentro de las tasas de mortalidad por edad lo constituyen las muertes de los menores de un año. El riesgo de morir es considerablemente más alto en el primer año de vida que en las edades siguientes, salvo las edades muy avanzadas.

Es precisamente en esta edad cuando la mortalidad es más sensible a los efectos del ambiente y si las tasas son altas una buena proporción de estas defunciones son evitables. Por ello esta medida es un indicador usual del nivel de salud e interesa particularmente conocerla.

Tasa de mortalidad infantil =

$$= \frac{\text{Defunciones de niños menores de 1 año en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Tal como en la tasa bruta de mortalidad y las tasas de mortalidad por sexo y edad, la constante que se utiliza en esta tasa es 1.000.

El numerador de la tasa de mortalidad infantil incluye las defunciones de ambos sexos y por todas las causas que ocurren dentro de un año calendario y en un área determinada en los niños que aún no han cumplido su primer año de vida. Dada la naturaleza de su numerador la tasa de mortalidad infantil tiene el carácter de una tasa de mortalidad por edad. Por lo tanto, debería esperarse que su denominador fuera la población de menores de 1 año de edad, estimada a mitad del mismo año calendario a que se refieren las muertes. Sin embargo, hay razones metodológicas por las cuales se hace necesario el uso de otro denominador.

Entre estas razones está el hecho de que la población menor de 1 año se omite en los censos en una proporción mayor, que la de cualquiera otra edad, y por ello su tamaño, para un año censal y con mayor razón en las estimaciones para los años posteriores al censo, son más inexactas que para los grupos de edades mayores. Por otra parte, los niños menores de 1 año que existen en una población depende del nivel y las tendencias de la natalidad en los años recientes. En cambio, en los grupos de edades mayores los efectivos de población son menos sensibles a las modificaciones de la natalidad en los años inmediatamente precedentes.

Es por esto, que para estar a cubierto de las variaciones que existen entre los países respecto a la cobalidad de los censos y de las fluctuaciones que puede experimentar el nivel de la natalidad, se ha convenido internacionalmente en utilizar como denominador de la tasa de mortalidad infantil la cifra de nacidos vivos del año, en lugar de la población estimada de menores de 1 año.

La tasa de mortalidad infantil se subdivide en dos componentes:

Tasa de mortalidad neonatal =

$$= \frac{\text{Defunciones niños menores de 28 ds. en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Esta tasa mide la frecuencia de muertes que ocurren en los menores de 28 ds. en un año calendario y en un área determinada por cada 1.000 nacidos vivos en ese mismo año y área.

Tasa de mortalidad infantil tardía =

$$= \frac{\text{Defunciones de niños de 28 ds. a 11 meses en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

La tasa de mortalidad infantil tardía mide la frecuencia de muertes que ocurren en el primer año de vida a partir del 28º día, en un año calendario y área dada por 1.000 nacidos vivos en ese año y área.

Así como entre las muertes del primer año es conveniente distinguir las que ocurren en las primeras 4 semanas del resto de las muertes infantiles, también es útil analizar separadamente las muertes de la primera semana de vida de las correspondientes a las 3 semanas siguientes. Si se refieren estos nuevos dos subconjuntos a la misma cifra de nacidos vivos del año se obtienen dos nuevas tasas que sumadas equivalen a la tasa de mortalidad neonatal. Ambas se expresan igualmente por 1.000. La tasa de mortalidad de la primera semana se denomina tasa de mortalidad neonatal precoz y la de la segunda a cuarta semana tasa de mortalidad neonatal tardía.

Tasa de mortalidad neonatal precoz =

$$= \frac{\text{Defunciones de menores de 7 días en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Esta tasa mide la frecuencia de muertes que ocurren en la primera semana de vida en un año calendario y área dada por cada 1.000 nacidos vivos del mismo año y área.

Tasa de mortalidad neonatal tardía =

$$= \frac{\text{Defunciones de niños de 7 a 27 días en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Mide la frecuencia de muertes que ocurren entre la segunda y cuarta semana de vida en un año calendario y área dada por cada 1.000 nacidos vivos del mismo año y área.

Tasa de mortalidad fetal tardía (o mortinatalidad) =

$$= \frac{\text{Defunciones fetales tardías (28 y + semanas de gestación) en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Según el momento de la gestación en que se produce la muerte del producto de la concepción, las defunciones fetales se clasifican en precoces (menos de 20 semanas de gestación) intermedias (20 a 27 semanas) y tardías (28 y más semanas de gestación). Las defunciones fetales tardías corresponden a los mortinatos y las precoces e intermedias a los abortos.

El registro de las defunciones fetales tiene una omisión importante. Esta omisión afecta principalmente a las defunciones fetales precoces.

Para las defunciones fetales tardías en cambio, el registro proporciona una información más completa, aunque siempre subestima la magnitud real del problema. Su denominador también son los nacidos vivos por las razones expuestas en la tasa de mortalidad materna.

Tasa de mortalidad perinatal =

$$= \frac{\text{Defunciones fetales tardías + defunciones de niños menores de 7 días en un área y año determinados}}{\text{Nacidos vivos en ese año y área}} \times 1.000$$

Esta tasa mide el riesgo de muerte que implica para el producto de la concepción el paso de la vida intrauterina a la vida extrauterina.

### 3.3 Medición de la morbilidad

El estudio de la morbilidad tiene serias dificultades. Desde luego, a diferencia de la muerte que ocurre una sola vez y en un momento bien definido y es un hecho permanente, la enfermedad puede ocurrir varias veces en la vida de un individuo, ya que se trata de una misma enfermedad o de enfermedades distintas y por último ellas pueden tener duración variable.

En lo que se refiere a la medición de la enfermedad se pueden distinguir tres tipos de unidades:

1) personas enfermas, 2) enfermedades, 3) episodios de enfermedad.

Por ejemplo, si una persona tiene durante el año 2 resfrios y 3 episodios diarreicos, se contabilizará:

a) persona enferma; b) 2 enfermedades; c) 5 episodios.

Por este motivo el Comité de Expertos en Estadísticas de Salud recomienda que en las estadísticas de morbilidad se especifique claramente a cual de estos tres criterios se refieren los datos.

En la medición de la morbilidad interesa fundamentalmente medir la frecuencia de la enfermedad en la población, su duración y su gravedad.

#### 3.3.a. Medición de la frecuencia de la enfermedad

Se distinguen dos tipos: la incidencia y la prevalencia.

##### Tasa de incidencia

Se denomina incidencia al número de casos nuevos que se presentan en un período de tiempo. Se refiere a enfermedades que comienzan durante un período definido y la tasa mide la frecuencia de acontecimientos que ocurren durante el período. En la tasa de incidencia se incluyen en el numerador los casos nuevos (enfermedades o enfermos) registrados durante el período y el denominador se refiere a la población estimada en el punto medio del período. Las tasas de incidencia pueden ser anuales pero también pueden referirse a cualquiera otra unidad de tiempo.



Tasa de incidencia =

$$= \frac{\text{Número de casos nuevos en el período}}{\text{Población a mitad del período}} \times 100.000$$

La tasa de incidencia muestra la dinámica de la enfermedad y expresa el riesgo de enfermar que tiene la población durante el período observado.

-Tasa de prevalencia

Prevalencia : es el número de casos (nuevos y antiguos) que se registran en un tiempo o momento dado, por ejemplo, el primer día de un mes o el último día de un año o el promedio diario dentro de un período de tiempo. La tasa de prevalencia tiene como numerador el número de casos que están presentes en ese momento y como denominador la población estimada para el mismo momento.

Tasa de prevalencia =

$$= \frac{\text{Número de casos existentes en un momento dado}}{\text{Población en ese momento}} \times 100.000$$

La tasa de prevalencia es una medida relativa cuyo sentido es comparable a la información que proporcionan los censos de población y mide sólo lo que existe o prevalece en ese momento. Es necesario hacer notar que en el numerador figuran todos los casos tanto los que se iniciaron antes del momento de medición como los casos nuevos que aparecen en ese momento.

Tratándose de enfermedades crónicas la prevalencia refleja mejor que la incidencia la magnitud del problema en la comunidad.

### 3.3.b. Medición de la gravedad de la enfermedad

Un aspecto de la morbilidad cuyo conocimiento tiene gran interés es la gravedad de la enfermedad. Ella puede medirse en términos de la incapacidad que produce. Por ejemplo, una enfermedad menor es aquella que no es causa de ausencia del trabajo. Esto hace necesario tener una escala de incapacidad para medir la severidad del cuadro. Además la medición tiene el problema de que la gravedad depende no sólo de la enfermedad sino que también de las características de los individuos que la padecen. Por ejemplo, un resfrío común puede ser motivo para que una persona guarde cama, mientras otro individuo con un resfrío de iguales condiciones continúa desarrollando sus actividades.

Por estas dificultades el índice de gravedad de una enfermedad que más se utiliza es la tasa de letalidad, que establece la relación entre los fallecidos por una enfermedad y los enfermos que padecen esa enfermedad.

Tasa de letalidad =

$$= \frac{\text{Número de defunciones por una enfermedad dada}}{\text{Número de enfermos de esa enfermedad}} \times 100$$

Mide la frecuencia con que se produce la muerte en una enfermedad. Esta es la tasa que permite establecer el pronóstico de las enfermedades.

### 3.3.c. Medición de la duración de la enfermedad

La duración de la enfermedad es un dato que interesa medir, entre otras razones, porque la enfermedad de mayor duración significa mayor costo. Puede hacerse esta medición en forma de un promedio. Por ejemplo, 60 enfermos de tifoidea estuvieron en cama un total de 1.080 días, la duración de la enfermedad es entonces:

$$\text{Duración} = \frac{1.080}{60} = 18 \text{ días en promedio}$$

Para la medición de la duración es necesario definir previamente que se entiende por enfermedad. En este caso la duración se refiere al tiempo promedio de estada en cama de los enfermos. Otras definiciones podrían tomar en cuenta, por ejemplo, el día de los primeros síntomas o el día en que se hizo el diagnóstico, etc.

El promedio puede obtenerse no solo en relación a los enfermos (60 en el ejemplo anterior) sino que puede obtenerse para episodios de enfermedad. Por ejemplo, en una escuela se registraron los resfriados de los alumnos y se tuvo un total de 100 resfriados en el año. La duración total de los resfriados fue de 500 días. La duración media de cada episodio fue, por lo tanto de 5 días.

### 3.4. Medición de la fecundidad

La medición de la fecundidad se hace a través de diferentes tipos de tasas que tratan de medir los niveles del fenómeno en un área.

#### Tasa bruta de natalidad

Es una tasa simple que relaciona los nacidos vivos registrados en un área geográfica durante un año con la población total de esta área.

Tasa bruta de natalidad =

$$= \frac{\text{Nacidos vivos en un área y año determinados}}{\text{Población total al 30/VI en ese año y área}} \times 1.000$$

Como incluye a la población total (de todas las edades y de ambos sexos) no puede interpretarse como una probabilidad porque en el denominador hay población que no está expuesta al riesgo de tener un niño. Expresa más bien la frecuencia de los nacimientos por cada 1.000 habitantes.

Las tasas de natalidad son prácticamente las únicas medidas de fecundidad que es posible calcular para áreas geográficas pequeñas y permite estudiar las tendencias del fenómeno en un área determinada.

Quando se comparan áreas diferentes hay que ser extremadamente cuidadoso en la interpretación porque puede haber diferencias en la estructura de la población especialmente en lo que se refiere a la composición por edad de la población femenina y esta diferencia puede por sí sola determinar diferencias en las tasas de natalidad.

#### Tasa de fecundidad general

Es esta una tasa más específica ya que tiene un denominador la población potencialmente expuesta al riesgo de tener un nacido vivo: la población femenina en edad fértil.

Tasa de fecundidad general =

$$= \frac{\text{Nacidos vivos en un área y año determinados}}{\text{Población femenina de 15 a 49 años al 30/VI en ese año y área.}} \times 1.000$$

Al tomar en cuenta solamente a las mujeres y en el grupo de edad expuesto al riesgo es una tasa más útil para hacer comparaciones entre zonas o comparaciones internacionales.

#### Tasa de fecundidad por edad

Esta tasa tiene un nuevo refinamiento y es más específica ya que toma en cuenta no solo el sexo, sino la composición por edad. En efecto, en su numerador se anotan los nacimientos de madres de un grupo de edad determinada y en el denominador la población femenina de esa edad.

Tasa de fecundidad por edad =

$$= \frac{\text{Nacidos vivos de mujeres de un grupo de edad en un área y año determinados}}{\text{Población femenina de ese grupo de edad al 30/VI en ese año y área.}} \times 1.000$$

Ej.: 
$$\frac{\text{Nacidos vivos de mujeres de 15 a 19 años en un área y año determinados}}{\text{Población femenina de 15 a 19 años al 30/VI en ese año y área.}} \times 1.000$$

Por lo general las tasas de fecundidad por edad se calculan para grupos quinquenales de edades comprendidas entre los 15 y los 49 años, es decir, se calculan 7 tasas de fecundidad por edad.

#### Otras medidas de fecundidad

Los estudios demográficos más finos de la fecundidad utilizan además de las tasas anteriores, las llamadas tasas de reproducción que tratan de medir el aporte futuro de la fecundidad al reemplazo de la población haciendo una corrección en los nacimientos utilizando la proporción de nacimientos femeninos.

Como se trata de tasas usadas por especialistas remitimos al lector a los textos de Demografía para su estudio.

V. ESTADISTICA DESCRIPTIVA

1.- Medidas de Posición

Tienen por objeto la obtención de un valor que resuma en sí todas las mediciones. La mayoría de ellas tratan de ubicar el centro de la distribución.

La expresión numérica de estas medidas es siempre un valor de la variable comprendido en su recorrido.

a) Media Aritmética

Es la medida de posición más conocida y utilizada, para su cálculo se hacen las siguientes consideraciones :

Media para datos no agrupados

Se calcula sumando todas las observaciones  $X_1, X_2, \dots, X_n$  y se divide por el número total de ellas y se denota por  $\bar{X}$  o  $M[X]$  según convenga.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = M[X]$$

Ejemplo : Encontrar la media de los siguientes datos : 55, 54, 51, 55, 53, 54, 52

Usando la calculadora Casio fx - 120 se tiene :

"SD". [INV] [AC] 55 [m+] 54 [m+] 51 [m+] 55 [m+] 53 [m+] [m+] 54 [m+] 52 [m+] [X] = 53,375

Media datos agrupados

En este caso las n observaciones están agrupadas en m intervalos de clases con marcas  $Y_1, Y_2, \dots, Y_m$  y funciones absolutas  $f_1, f_2, \dots, f_m$  luego el valor de la media se calcula por

$$\bar{X} = \frac{\sum_{i=1}^m f_i Y_i}{\sum_{i=1}^m f_i} = \frac{\sum_{i=1}^m f_i Y_i}{n}$$

Ejemplo : En la tabla construida anteriormente en la cual se agrupan los pesos de 40 estudiantes se tiene que el peso promedio es :

$$\bar{X} = \frac{\sum_{i=1}^{12} f_i y_i}{40} = 147,25 \text{ lb.}$$

### Media de Submuestras ( o subpoblaciones )

Cuando existen varios grupos o submuestras las cuales forman una partición de una muestra global ( es decir no tienen elementos comunes y en conjunto forman la muestra global ). De estos grupos se conoce su tamaño  $n_i$  y su media  $\bar{Y}_i$  y se desea determinar la media de la muestra global. Esta se calcula de la siguiente forma

$$\bar{Y} = \frac{\sum_{i=1}^p n_i \bar{Y}_i}{n} \quad ; \quad \begin{aligned} p &= \text{N}^\circ \text{ de grupos} \\ n &= \text{Tamaño muestral global} \\ &= \sum_{i=1}^p n_i \end{aligned}$$

Ejemplo : En un colegio de 65% de los funcionarios son Profesores el 15% Inspectores, el 15% Administrativos y el 5% Auxiliares. Los Profesores tienen una edad promedio de 45 años, los Inspectores de 28 años, los Administrativos de 35 años y los Auxiliares de 25 años. ¿Cuál es la edad promedio de los funcionarios ?

Respuesta

$$\begin{aligned} \bar{X} &= \frac{0,65 n 45 + 0,15 n 28 + 0,15 n 35 + 0,05 n 25}{n} \\ &= 0,65 45 + 0,15 28 + 0,15 35 + 0,05 25 \\ &= 39,95 \end{aligned}$$

### Propiedades de la Media

- i)  $\overline{aX + k} = a\bar{X} + k$
- ii)  $\overline{k} = k$
- iii)  $\overline{X - \bar{X}} = 0$



b) Mediana

La Mediana es una medida que divide la distribución en dos grupos con igual número de observaciones.

Una vez ordenadas las observaciones en orden creciente o decreciente, llamaremos mediana, que denotaremos por Me al valor de la variable que supere a lo sumo a la mitad de las observaciones y sea superado a lo sumo por la otra mitad.

Para su cálculo pueden presentarse dos casos :

i) Datos no agrupados

- Ordenar los datos según magnitud
- Determinar la posición correspondiente a la mediana es decir la posición  $\frac{n+1}{2}$  ; en que "n" representa el N° de observaciones.
- Ubicar el valor correspondiente a la  $\frac{n+1}{2}$  observación

Nota Cuando el número de observaciones es par; no hay una observación central sino dos: en este caso se adopta el criterio de definir la Mediana como el promedio de estas dos observaciones centrales.

Ejemplo Dados los siguientes datos : 18, 14, 16, 17, 19, 22 ordenándolos se tiene 14, 16, 17, 18, 19, 22.

como hay seis datos; tenemos dos observaciones centrales 17 y

$$18; \text{ luego } Me = \frac{17+18}{2} = 17.5$$

ii) Datos agrupados

Cuando los datos están agrupados, no es posible individualizar rápidamente la mediana y sólo es factible determinar el intervalo en que se encuentra.

Para esto se empieza por establecer la menor de las frecuencias absolutas acumuladas que supere a la mitad del total de observaciones; es decir es necesario encontrar un valor j para

el cual  $F_j > \frac{n}{2}$  y luego aplicar la siguiente fórmula para

$$\text{ra Me : } Me = Y'_{j-1} + \left[ \frac{c_j \left( \frac{n}{2} - F_{j-1} \right)}{f_j} \right]$$

$Y'_{j-1}$  = límite inferior del intervalo j-ésimo

$F_{j-1}$  = Frecuencia absoluta acumulada del intervalo anterior al  $j$ -ésimo

$c_j$  = amplitud o anchura del intervalo  $j$ -ésimo

Ejemplo : En el ejemplo donde se presentaba el peso de 40 estudiantes se tiene :  $\frac{n}{2} = 20$ , buscando en la columna de frecuencias absolutas acumuladas vemos que  $F_j = 23 > 20$

luego  $j = 6$

Por lo tanto la mediana está en el intervalo

143 - 148 ; luego

$$\begin{aligned} \text{Me} &= 143 + \frac{5(20 - 15)}{8} \\ &= 143 + \frac{25}{8} = 146,125 \end{aligned}$$

### c) Moda

Es un estadígrafo de posición que puede definirse como el valor observado más frecuente; es decir; el que presentó una frecuencia mayor.

Se le emplea de preferencia a otros estadígrafos cuando se desea señalar el más común o típico y se le denota como  $M_d$ .

En una distribución de frecuencias con intervalos iguales, la Moda puede considerarse como la marca de clase del intervalo Modal, es decir; el intervalo de mayor frecuencia.

Por ejemplo en el caso de los pesos de 40 estudiantes se tiene que  $M_d = 145,5$

si se desea precisar el valor de la moda para datos agrupados en distribuciones de frecuencia se pueden utilizar las siguientes fórmulas :

$$\begin{aligned} M_d &= Y'_{j-1} + C_j \left[ \frac{f_{j+1}}{f_{j-1} + f_{j+1}} \right] \\ M_d &= Y'_{j-1} + C_j \left[ \frac{f_j - f_{j-1}}{(f_j - f_{j-1}) + (f_j - f_{j+1})} \right] \\ &= \text{lim inf } (j) + C_j \left[ \frac{M_1}{M_1 + M_2} \right] \quad \begin{array}{l} M_1 = f_j - f_{j-1} \\ M_2 = f_j - f_{j+1} \end{array} \end{aligned}$$

c) Media geométrica

Se usa especialmente en los casos en que existe una tasa de crecimiento relativamente constante. Se usa también de preferencia cuando es conveniente dar importancia a los valores pequeños

i) Datos no agrupados

La media geométrica de un conjunto de  $n$  valores positivos se define como la raíz  $n$ -ésima del producto de los  $n$  valores y se denota por  $M_g$ . luego  $M_g = \sqrt[n]{X_1 X_2 \dots X_n}$  ✓

ii) Datos agrupados

Si se tienen  $n$  datos agrupados en  $m$  clases o intervalos con marcas de clase  $Y_1, Y_2 \dots Y_m$  y frecuencias absolutas  $f_1, f_2 \dots, f_m$  respectivamente, entonces la media geométrica está dada por:

$$M_g = \sqrt[n]{\frac{f_1}{y_1} \frac{f_2}{f_2} \dots Y_m}$$

Como el cálculo de la expresión anterior es muy largo se adopta el siguiente criterio :

aplicando logaritmos a ambos lados se obtiene :

$$\log M_g = \frac{\sum_{i=1}^m f_i \log Y_i}{n} = M [\log Y]$$

es decir el logaritmo de la media geométrica de una variable es la media aritmética del logaritmo de esa variable.

Ejemplo En el ejemplo de los pesos de 40 estudiantes se tiene:

$$\log M_g = \frac{86,668}{40} = 2,1667$$

$$M_g = 146,79$$

e) Media Armónica

La Media Armónica denotada por  $M_h$  se define como el valor recíproco de la Media Aritmética de los valores recíprocos de la variable. Es decir.



$$M_h = \frac{1}{M \left[ \frac{1}{x} \right]} = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

### Datos Agrupados

Cuando se tienen  $n$  observaciones agrupadas en  $m$  clases o intervalos con marcas de clase  $Y_1, Y_2, \dots, Y_m$  y frecuencias absolutas

$f_1, f_2, \dots, f_m$  respectivamente entonces

$$M_h = \frac{1}{M \left[ \frac{1}{y} \right]} = \frac{1}{\sum_{i=1}^m f_i \frac{1}{Y_i}} = \frac{n}{\sum_{i=1}^m \frac{f_i}{Y_i}}$$

Ejemplo : Para el peso de los 40 estudiantes

$$M_h = \frac{40}{0,2735} = 146,25$$

### f) Cuartiles, deciles y percentiles

Los cuartiles son valores que dividen una serie de datos en cuatro partes iguales, dejando el 25% de los casos en cada una. Los cuartiles son 3 y se denotan por  $Q_i$ ,  $i = 1, 2, 3$ .

Los deciles son valores que dividen la serie de datos en 10 partes iguales, dejando el 10% de los casos en c/u. Los deciles son 9 y se denotan por  $D_i$ . Los percentiles, dividen la serie de datos en 100 partes, dejando el 1% de los casos en c/u. Los percentiles son 99 y se denotan por  $P_i$ .

### Observaciones

i) Sabemos que la mediana divide la serie de datos en dos partes dejando el 50% de los casos en c/u; de modo que también podemos llamarle percentil 50 ( $P_{50}$ ); decil 5 ( $D_5$ ) o cuartil 2

( $Q_2$ )

ii) Se puede ver que los cuartiles y los deciles son un caso particular de percentiles; por lo que sólo veremos el cálculo de éstos y para calcular deciles y cuartiles haremos la analogía correspondiente

### - Cálculo de percentiles para datos no agrupados

Si tenemos una serie de "n" datos ordenados por magnitud entonces el valor del percentil r-ésimo ( $r = 1, 2, \dots, 99$ ) está por la observación o dato que ocupa el lugar  $\frac{(n+1) \cdot r}{100}$ ;

en caso que  $\frac{(n+1) \cdot r}{100}$  no sea un número entero, se tendrá que interpolar.

Ejemplo : Dada la siguiente serie de datos ordenados 42, 42, 43, 45, 47, 50, 53 y necesitamos calcular el percentil 67 ( $P_{67}$ )

Calculamos el valor  $\frac{(n+1) \cdot r}{100} = \frac{8 \cdot 67}{100} = \frac{536}{100} = 5,36$

luego el percentil 67 se encuentra en la 5<sup>ta.</sup> y la 6<sup>ta.</sup> observación; luego deberos interpolar  $P_{67}$  está entre 47 y 50

luego  $\frac{47}{0,36} \quad P_{67} \quad \frac{50}{0,64}$

$$\frac{P_{67} - 47}{0,36} = \frac{50 - P_{67}}{0,64}$$

$$P_{67} = 0,36 \cdot 50 + 0,64 \cdot 47 = 48,08$$

### Cálculo de Percentiles para datos agrupados

Si tenemos una serie de n datos agrupados en m clases o intervalos y necesitamos calcular el percentil r ( $P_r$ ). Para ello es necesario primero identificar el intervalo que lo contiene y ésto se logra encontrando la menor de las frecuencias absolutas acumuladas que supere el valor  $\frac{n \cdot r}{100}$

Es decir hay que hallar el intervalo j-ésimo tal que

$F_j > \frac{n \cdot r}{100}$  , luego el valor de  $P_r$  se pa

$$P_r = Y'_{j-1} + \left[ \frac{C_j \left( \frac{n \cdot r}{100} - F_{j-1} \right)}{f_j} \right]$$

$r = 1, 2, \dots, 99$

Ejemplo : Para el caso del peso de los 40 estudiantes se desea calcular el peso mínimo del 20 % superior del total de estudiantes, es decir  $P_{80}$

$$\frac{n \cdot r}{100} = \frac{40 \cdot 80}{100} = 32 \implies j=9$$

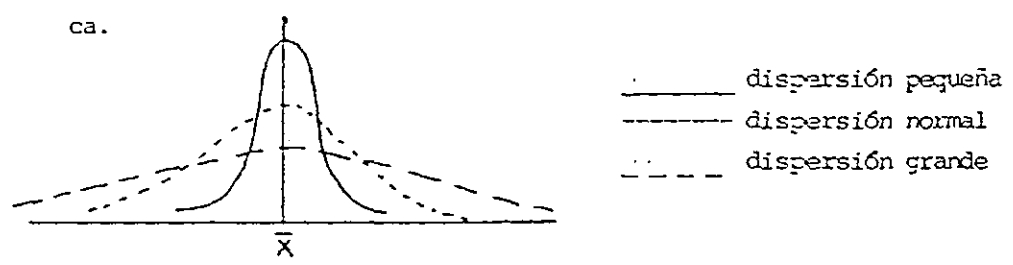
y el intervalo j-ésimo es [158 - 163)

$$\text{luego } P_{80} = 158 + \frac{5(32 - 32)}{2} = 158$$

Luego el 20 % superior de los estudiantes tiene un peso mínimo de 158 libras o el 80% inferior de los estudiantes tiene un peso máximo de 158 libras.

2.- Medidas de dispersión

La idea de dispersión se relaciona con la mayor o menor concentración de datos en Torno a un valor central, generalmente la media aritmética.



a continuación examinaremos algunas medidas que pretenden cuantificar esta dispersión en una distribución de frecuencias:

a) Varianza

La varianza de una distribución de frecuencia de una variable X es la media aritmética de los cuadrados de las diferencias entre los valores de la distribución y su media aritmética. Se denota por  $s_x^2$  o  $V(X)$  y se calcula de la siguiente forma:

$$s_x^2 = V(X) = \begin{cases} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} & \text{para datos no agrupados} \\ \frac{\sum_{i=1}^m f_i (y_i - \bar{X})^2}{n} & \text{para datos agrupados} \end{cases}$$

Ejemplo : Para el caso del peso de 40 estudiantes

$y_i$	$f_i$	$y_i - Y$	$(y_i - Y)^2$	$f_i (y_i - Y)^2$

$$\text{luego } S_x^2 = \frac{\sum_{i=1}^{12} f_i (y_i - \bar{Y})^2}{40} =$$

b) Desviación Estandar

Se define como la raíz cuadrada positiva de la varianza y se denota por  $S_x$  y se mide en las mismas unidades que la variable.

$$\text{Luego : } S_x = + \sqrt{S_x^2} = + \sqrt{V(X)}$$

Propiedades de la Varianza

$$\begin{aligned} \text{i) } S_x^2 &= \frac{\sum_{i=1}^m f_i (y_i - \bar{X})^2}{n} \\ &= \frac{\sum_{i=1}^m f_i (y_i^2 - 2y_i\bar{X} + \bar{X}^2)}{n} \\ &= \frac{\sum_{i=1}^m f_i y_i^2}{n} - 2\bar{X} \frac{\sum_{i=1}^m f_i y_i}{n} + \bar{X}^2 \frac{\sum_{i=1}^m f_i}{n} \\ &= \frac{\sum_{i=1}^m f_i y_i^2}{n} - \bar{X}^2 \\ &= \overline{y^2} - \bar{X}^2 \end{aligned}$$

ii)  $V(k) = 0$ ; k cte.

iii)  $V(ax + k) = a^2 V(X)$

c) Coefficiente de Variación

Se utiliza cuando las variables en dos distribuciones se expresan en unidades diferentes. No tiene sentido comparar algunos de los estadígrafos de dispersión presentados, pues ellos quedan influenciados por el valor numérico de dichas unidades.

Por ejemplo una desviación estandar de 2,7 kilos ¿ Es mayor que una de 0,48 años ? . Esta dificultad se puede superar mediante el

coeficiente de variación, que no depende del sistema de unidades empleados y se calcula como el cociente entre la desviación estándar y la media aritmética y se denota por :

$$C.V. = \frac{S}{\bar{X}}$$

3.- Tablas bidimensionales

Anteriormente se ha considerado el caso de distribuciones de valores de una sola variable. A menudo se presentan las observaciones respecto a dos variables en forma simultánea. Por ejemplo peso y edad de un grupo de estudiantes, estatura de padre e hijo, ingreso y gastos por familia, etc., en este caso se habla de distribuciones bidimensionales las cuales también es posible agruparlos en tablas de distribución de frecuencias conjuntas, cuando el N° de observaciones es muy grande, e incluso es posible averiguar si existe alguna relación funcional entre ambas variables.

a) Presentación de los datos

Si el N° de observaciones es reducido los datos pueden presentarse en forma simple por ejemplo : Salario y gasto en alimentación en 6 familias

SALARIO \$ (x)	73	65	69	76	83	47
GASTOS \$ (y)	48	40	48	43	53	33

cuando los datos son más numerosos se pueden agrupar en una tabla de doble entrada o tabla de distribución de frecuencias conjunta.

Para confeccionar este tipo de tabla es necesario determinar los intervalos de cada variable por separado

Por ejemplo : Edad (y) Y N° de caries (X) con un grupo de 30 niños.

	2	4	6	8	10	
1	3	5	7	9		
2	4	2			6	
3	3	6	1		10	
4		1	5		6	
5			3	2	5	
6			1	2	3	
	7	9	10	4	30	

En la última columna se observa las frecuencias absolutas para X y la última fila las frecuencias absolutas para Y y dentro de la tabla se tienen las frecuencias conjuntas y denotaremos por  $f_{ij}$  y corresponde a la frecuencia del  $i$ -ésimo intervalo para X y  $j$ -ésimo intervalo para Y. En esta tabla podemos calcular:  $\bar{X}, \bar{Y}, N(X), N(Y) \in [4,8), P_{20}, P_{80}$ .

b) Covarianza

Es una medida que cuantifica la relación entre dos variables. Representa el grado de variabilidad conjunta y se define como la media de los productos de las desviaciones respecto a sus correspondientes medias aritméticas y se denota por  $Cov(X, Y)$

i) Datos no agrupados

$$Cov(X, Y) = \frac{\sum_{i=1}^m (x_i - \bar{X})(y_i - \bar{Y})}{n}$$

ii) Datos agrupados

$$Cov(X, Y) = \frac{\sum_{i=1}^m \sum_{j=1}^l (x_i - \bar{X})(y_j - \bar{Y}) f_{ij}}{n}$$

donde  $m = N^\circ$  de intervalos para X

$l = N^\circ$  de intervalos para Y

$x_i =$  marca de clase del intervalo  $i$ -ésimo para X

$y_j =$  marca de clase del intervalo  $j$ -ésimo para Y

Nota : Para efectos de cálculo de la covarianza entre X e Y se utiliza la siguiente fórmula que se deduce fácilmente de la fórmula anterior.

$$Cov(x, y) = \overline{XY} - \bar{X} \cdot \bar{Y}$$

Ejemplo : Calcular la covarianza en el ejemplo anterior :

$$Cov(X, Y) = \overline{XY} - \bar{X} \cdot \bar{Y} = \frac{\sum_{i=1}^5 \sum_{j=1}^4 f_{ij} X_i Y_j}{n} - \frac{\sum_{i=1}^5 f_{i.} X_i}{n} \cdot \frac{\sum_{j=1}^4 f_{.j} Y_j}{n}$$

$$= 34,3 - 5,27 \cdot 5,73 = 4,1029$$

### 2.- Ajuste de modelos por mínimos cuadrados

Al graficar dos variables, la nube o conjunto de puntos adopta diferentes formas, las que se pueden esquematizar en una línea más sencilla y proporcionar así una idea del tipo de relación funcional entre ambas variables. Esta relación funcional permite la posibilidad de predecir (o estimar) el valor de una variable dado el valor de la otra.

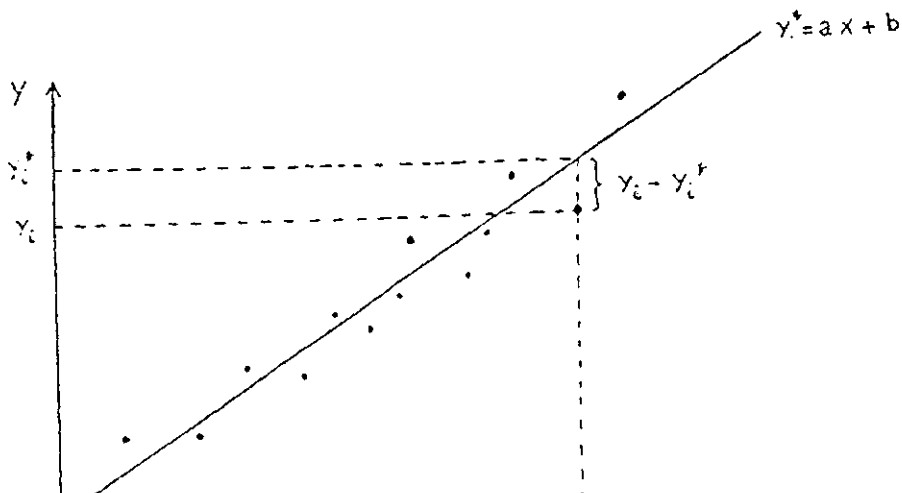
El problema es, entonces, trazar través de la nube de puntos la línea que mejor ajuste los datos, para hacer las predicciones.

Como podemos trazar infinitas líneas la búsqueda debe ser única y óptima en el sentido de que el error de estimación o diferencia entre el valor observado  $Y_j$  y el estimado  $Y_j^*$  sea lo más reducido posible (sea mínimo).

El mejor modelo será aquel que minimice la suma de los cuadrados del error de estimación

$$\sum_{i=1}^n (Y_i - Y_i^*)^2$$

La función más simple utilizada como ajuste es la de primer grado en X cuya expresión analítica es  $Y^* = ax + b$  y que se denomina recta de regresión de Y en X, donde X es la variable independiente (toma cualquier valor) e Y es la variable dependiente (depende del valor que tome X).



Los parámetros son la pendiente "a" y la ordenada en el origen "b".  
Los valores de a y b que minimizan la función

$$F(a,b) = \sum_{i=1}^n (y_i - Y_i^*)^2 \quad \text{son}$$

$$\hat{a} = \frac{\text{cov}(x,y)}{S_x^2} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - \bar{X}^2}$$

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

La recta de regresión de X e Y es  $X^* = CY + d$  y por el mismo método

podemos hallar que :  $C = \frac{\text{cov}(x,y)}{S_y^2}$  y  $d = \bar{X} - C\bar{Y}$

Ejemplo Dada la siguiente tabla en la cual se muestran el ingreso (x) de 10 familias y los gastos en alimentación (y)

x	73	65	69	76	83	47	70	58	65	70
y	48	40	48	43	53	33	43	39	43	45

ajustar una recta para estimar los gastos en alimentación.

#### Solución

Como se desea estimar los gastos en alimentación (variable y), entonces ésta es la variable dependiente luego la ecuación de regresión es :

$$y^* = ax + b \quad \text{donde } \hat{a} = \frac{\text{cov}(x,y)}{S_x^2} = \frac{\overline{XY} - \bar{X}\bar{Y}}{S_x^2} = \frac{2985,1 - 2940,6}{88,04}$$

$$= \frac{44,5}{88,04} = 0,5055$$

$$\hat{b} = \bar{Y} - \hat{a}\bar{X} = 43,5 - 0,5055 \cdot 67,6 = 9,3282$$

el modelo es :  $Y^* = 0.505 x + 9.3282$

¿Cuál serían los gastos en alimentación para una familia cuyo ingreso es de 52 dólares ?

$$y^* = 0,5055 \cdot 52 + 9,3282 = 35,6 \text{ dólares}$$

#### Correlación lineal

Al ajustar una recta a un conjunto de puntos, a veces, los puntos se encuentran relativamente próximos a la línea y otras, quedan



bastante dispersos en torno a ella. Luego es necesario determinar una medida que cuantifique esta "aproximación"

Esta medida se denomina coeficiente de correlación lineal denotado por  $\rho_{xy}$  y está dado por :

$$\rho_{xy} = \frac{\text{cov}(x,y)}{s_x s_y} = \frac{\overline{XY} - \bar{X} \bar{Y}}{s_x s_y}$$

#### NOTAS

- i)  $-1 \leq \rho_{xy} \leq 1$
- ii) si  $\rho_{xy} > 0$  diremos que hay correlación directa
- iii) si  $\rho_{xy} < 0$  diremos que hay correlación inversa
- iv) si  $\rho_{xy} = 1$  diremos que hay correlación directa perfecta
- v) si  $\rho_{xy} = -1$  diremos que hay correlación inversa perfecta
- vi) si  $\rho_{xy} = 0$  diremos que no hay correlación lineal

Ejemplo El coeficiente de correlación lineal entre el ingreso y el gasto en alimentación en 10 familias es :

$$\rho_{xy} = \frac{44,5}{\sqrt{88,04 \cdot 27,65}} = 0,902$$

luego hay una correlación directa casi perfecta.

#### Modelo Exponencial

Es de la forma  $Y = a \cdot b^X$  y para ajustar este tipo de modelo, se reduce a un modelo lineal simple aplicando logaritmos y para hallar los parámetros se procede de igual forma que la anterior.

El modelo reducido queda :  $\log y = x \log b + \log a$ , que es de la forma  $y' = a'x + b'$  donde  $y' = \log y$ ,  $a' = \log b$ ,  $b' = \log a$

$$\text{luego } a' = \frac{\text{cov}(x,y)}{s_x^2} ; b' = \bar{y}' - a' \bar{x}$$

y para obtener a y b se aplican los correspondientes antilogaritmos

#### Modelo Potencial

Es de la forma  $y = ax^b$ , para obtener los parámetros a y b se procede de la siguiente forma  $\log y = \log a + b \log X$  haciendo  $y' = \log y$ ;  $b' = \log a$ ,  $a' = b$ ,  $x' = \log x$

donde  $a' = b = \frac{\text{cov}(x', y')}{S_{x'}}$ ,  $b' = \bar{Y}' - a' \bar{X}'$

¿Cómo saber cuál de estos modelos debemos ajustar a un conjunto de datos?

La respuesta la proporciona el coeficiente de correlación lineal entre las variables correspondientes  $\rho_{xy}$ ,  $\rho_{x \log y}$ ,  $\rho_{\log x, y}$ ,  $\rho_{\log x \log y}$  el valor de  $|\rho|$  más grande nos proporciona un "mejor" ajuste

	<u>Modelo a ajustar</u>
Para $\rho_{xy}$	$y = ax + b$ o $x = a'y + b'$
Para $\rho_{x \log y}$	$y = ab^x$
Para $\rho_{\log x, y}$	$x = ab^y$
Para $\rho_{\log x, \log y}$	$y = ax^b$ o $x = a'y^{b'}$

VI TEORÍA ELEMENTAL DE PROBABILIDADES

Definición 1 Sea  $\mathcal{E}$  un Experimento aleatorio. Llamaremos espacio muestral  $\Omega$  al conjunto de todos los resultados posibles del experimento.

Definición 2 Sea  $\Omega$  un espacio de muestras asociado a un experimento  $\mathcal{E}$ .

- i) Llamaremos suceso a cualquier subconjunto de  $\Omega$ .
- ii) Llamaremos suceso elemental a los sucesos que tengan solo un resultado posible
- iii) Diremos que dos sucesos A y B de  $\Omega$  son mutuamente excluyentes ssi. no tienen resultados comunes: es decir  $A \cap B = \emptyset$
- iv)  $\Omega$  y  $\emptyset$  son sucesos, llamados suceso seguro y suceso imposible respectivamente.

Observación

Si un experimento  $\mathcal{E}$  tiene como espacio de muestras asociado a  $\Omega$  cuando éste se realiza una vez, entonces  $\Omega \times \Omega \times \dots \times \Omega$  ( $n$  veces)  $(\Omega^n)$  es el espacio muestral cuando se realiza el exp.  $n$  veces.

Definición Clásica de Probabilidad

Si un suceso A puede ocurrir en h formas de un total de n formas posibles e igualmente probables, entonces la probabilidad de que ocurra el suceso A es  $P(A) = \frac{h}{n} = \frac{n \text{ casos favorables}}{n \text{ casos posibles}} = \frac{n \text{ elementos de A}}{n \text{ elementos de } \Omega}$

y la posibilidad de no ocurrencia de A es

$$\begin{aligned} P(A^c) &= \frac{n-h}{n} = \frac{n}{n} - \frac{h}{n} \\ &= 1 - P(A) \end{aligned}$$

Observación

i) Si A es un suceso entonces  $0 \leq P(A) \leq 1$

ii)  $P(A) + P(A^c) = 1$

Propiedades de P

i)  $P(\Omega) = 1$  y  $P(\emptyset) = 0$

ii) Si A y B son mutuamente excluyentes entonces  $P(A \cup B) = P(A) + P(B)$

iii)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

iv) si  $A \subset B \Rightarrow P(A) \leq P(B)$

v) si  $A \subset B \Rightarrow P(B-A) = P(B) - P(A)$

Probabilidad condicional

Definición Sea B un suceso dado tal que  $P(B) > 0$  entonces la probabilidad de un suceso A condicionado a la ocurrencia de B se define por

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{y se lee probabilidad de A dado B}$$

Definición Sea A y B sucesos, entonces diremos que A y B son independientes si  $P(A \cap B) = P(A) \cdot P(B)$

Observación

i)  $P(A \cap B) = P(B) \cdot P(A/B)$

ii) si A y B indep.  $P(A/B) = P(A)$  y  $P(B/A) = P(B)$

Ejemplo resuelto

En un estudio hecho en el Servicio de Selección y Registro de Estudiantes de la Universidad de Chile, sobre el número de postulantes a las Universidades Chilenas y el sexo, se obtuvo la siguiente información:

SEXO	POSTULA A LA UNIVERSIDAD		TOTAL
	SI	NO	
FEMENINO	10	12	22
MASCULINO	82	96	178
TOTAL	92	108	200

- a) Si se elige al azar un estudiante de esta muestra de 200 ¿Cuál es la probabilidad de que :
- i) Sea de sexo femenino
  - ii) No haya postulado a la Universidad
  - iii) No haya postulado y sea de 'sexo masculino
  - iv) Haya postulado a la Universidad y sea sexo femenino
  - v) Son independientes los sucesos ser postulante a la Universidad y ser de sexo masculino
  - vi) Si se eligen al azar dos estudiantes ¿Cuál es la probabilidad de que :
    - i) Al menos uno sea de sexo femenino
    - ii) los dos hayan postulado a la Universidad
    - iii) Ninguno sea de sexo femenino

Solución

a) Sean los sucesos

F : sexo femenino

P : postulante a la Universidad

M : sexo masculino

P' : no postulante

luego

$$i) P(F) = \frac{22}{200}$$

$$ii) P(P') = \frac{108}{200}$$

$$iii) P(P' \cap M) = \frac{96}{200}$$

$$iv) P(P \cap F) = \frac{10}{200}$$

$$v) P(P \cap M) = \frac{82}{200} ; P(P) = \frac{92}{200} ; P(M) = \frac{178}{200}$$

$$P(P) \cdot P(M) \approx 0,4094 \neq 0,41 = P(P \cap M)$$

Luego no son independientes

b) Sean los sucesos :

$F_i$  : el  $i$ -ésimo estudiante es de sexo femenino

$P_i$  : el  $i$ -ésimo estudiante postula a la Universidad

$$\begin{aligned} \text{i) } P(F_1 \cap F_2^C) + P(F_1^C \cap F_2) + P(F_1 \cap F_2) &= \\ &= P(F_1) P(F_2^C / F_1) + P(F_1^C) P(F_2 / F_1^C) + P(F_1) P(F_2 / F_1) \\ &= \frac{22}{200} \cdot \frac{178}{199} + \frac{178}{200} \cdot \frac{22}{199} + \frac{22}{200} \cdot \frac{21}{199} = 0,21 \end{aligned}$$

$$\text{ii) } P(P_1 \cap P_2) = P(P_1) P(P_2 / P_1) = \frac{92}{200} \cdot \frac{91}{199} = 0,21$$

$$\text{iii) } P(F_1^C \cap F_2^C) = P(F_1^C) P(F_2^C / F_1^C) = \frac{178}{200} \cdot \frac{177}{199} = 0,79$$

#### Ejercicios propuestos

1.- En un curso de  $M$  alumnos existen  $r$  ( $r < M$ ) alumnos con problemas de aprendizaje. Se les aplica una prueba uno por uno hasta encontrar uno con problemas.

Describa un espacio de Muestras para este experimento.

2.- Suponga que a los alumnos del problema 1 se les aplica la prueba hasta encontrar a todos los alumnos con problemas de aprendizaje. Describir un espacio de muestras para este experimento.

3.- Consideremos cuatro individuos  $a, b, c$  y  $d$ . Supongamos que el orden en el cual se colocan estos individuos re presenta un resultado del experimento.

Sean  $A$  y  $B$  los sucesos definidos como :

$A = \{ "a" \text{ está en primer lugar} \}$

$B = \{ "b" \text{ está en segundo lugar} \}$

i) Anote todos los elementos del espacio muestral

ii) Anote los elementos de los sucesos  $A \cap B$  y  $A \cup B$

4.- Suponga que  $A$  y  $B$  son sucesos para los cuales

$$P(A) = x; P(B) = y; P(A \cap B) = z$$

Expresar cada una de las probabilidades siguientes en términos de  $x, y, z$ .

i)  $P(A^C \cup B^C)$

iii)  $P(A^C \cup B)$

ii)  $P(A^C \cap B)$

iv)  $P(A^C \cap B^C)$

5.- En un curso existen 10 alumnos normales; cuatro con problemas de aprendizaje y dos con retardo.

i) Si se elige un alumno al azar. Encontrar la probabilidad de que :

- a) No tenga ningún problema
- b) Tenga retardo
- c) Sea normal o tenga un retardo

ii) Si se eligen dos alumnos sin reemplazo . Encuentre la probabilidad de que :

- a) Ambos sean normales
- b) A lo menos uno sea normal
- c) Exactamente uno sea normal
- d) Ninguno sea normal
- e) Ambos tengan retardo
- f) A lo mas uno sea normal
- g) Ninguno tenga retardo

6.- La siguiente tabla muestra la distribución de 400 estudiantes según aprobación en una asignatura y reforzamiento

RESULTADO	REFORZAMIENTO		TOTAL
	SI	NO	
APRUEBA	140	110	250
REPRUEBA	50	100	150
TOTAL	190	210	400

a) Si elegimos una persona al azar ¿Cuál es la probabilidad de que ? :

- i) Aprueba y haya tenido reforzamiento
- ii) No aprueba dado que tuvo reforzamiento
- iii) No tenga reforzamiento dado que aprueba
- iv) No apruebe o tenga reforzamiento

b) Los sucesos " aprobar " y " Tener reforzamiento " ¿ Son independientes ?

7.- Sean A y B dos aptitudes. Dado que estas aptitudes son independientes y que las probabilidades de que un estudiante presente las aptitudes A y B son 0,50 y 0,75 respectivamente.

- a) ¿Cuál es la probabilidad de que un estudiante presente ambas aptitudes ?
- b) ¿Cuál es la probabilidad de que un estudiante presente por lo menos una aptitud ?
- c) ¿Cuál es la probabilidad de que un estudiante no presente ninguna aptitud ?

## VII DISTRIBUCIONES DE PROBABILIDADES MAS USUALES

### 1.- Distribución Binomial

Si  $p$  es la probabilidad de ocurrencia de un suceso en un solo ensayo de un experimento y  $q = 1 - p$  es la probabilidad de no ocurrencia del suceso entonces la probabilidad de que el suceso ocurra exactamente  $k$  veces en  $n$  repeticiones independientes del experimento está dada por :

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}; k= 0, 1, 2, \dots, n$$

esta distribución de probabilidades discreta recibe el nombre de distribución binomial de parámetros  $n$  y  $p$

#### Requisitos para aplicar la distribución Binomial

- Debe haber un número fijo  $n$  de ensayos
- En cada ensayo los resultados posibles son sólo dos ; denominados " éxito " y " fracaso "
- La probabilidad de " éxito " debe ser igual en todo los ensayos
- Los ensayos deben ser independientes entre sí

#### Propiedades

$$\text{Medida o valor esperado} = n p$$

$$\text{desviación estandar} = \sqrt{np (1-p)}$$

#### Ejemplo resuelto

El fracaso escolar en un colegio es habitualmente de un 10%. Si se eligen dos estudiantes del colegio al azar ¿Cuál es la probabilidad de que :

- i) Uno de ellos fracase y el otro apruebe ?
- ii) Los dos fracasen ?
- iii) Los dos aprueben ?

Solución

Sea  $X = N^\circ$  de fracasos en dos estudiantes

con  $p = 0,10$  ;  $q = 1 - p = 0,90$

$$i) P(X=1) = \binom{2}{1} (0,10)^1 (0,90)^1 = 2 \cdot 0,90 = 0,18$$

es decir de dos estudiantes hay un 18% de posibilidades de que uno fracase.

$$ii) P(X=2) = \binom{2}{2} (0,10)^2 (0,90)^0 = 0,01$$

es decir, hay un 1% de posibilidades de que ambos fracasen

$$iii) P(X=0) = \binom{2}{0} (0,10)^0 (0,90)^2 = 0,81$$

es decir hay un 81% de posibilidades de que ambos aprueben.

Si en el colegio se matriculan 200 alumnos

¿Cuál será el  $N^\circ$  esperado de fracaso escolar ?

$$N^\circ \text{ esperado} = n p = 200 (0,1) = 20$$

se espera que fracasen 20 alumnos

Ejercicios Propuestos

1.- Si el 20% de los estudiantes de un colegio presenta problemas de aprendizaje

a) Determinar la probabilidad de que dé 4 estudiantes elegidos al azar de este colegio :

- i) Uno tenga problemas de aprendizaje
- ii) Ninguno tenga problemas de aprendizaje
- iii) A lo más dos tengan problemas de aprendizaje

b) Si en un año determinado se matriculan 400 estudiantes en este colegio

¿Cuál será el número esperado de estudiantes con problemas de aprendizaje ?

2.- En una asignatura determinada generalmente el 10% de los estudiantes reprobaban. Si se inscriben 20 estudiantes en esa asignatura:

a) ¿Cuál es la probabilidad de que ?:

- i) cinco estudiantes reprobaban
- ii) a lo más 3 estudiantes reprobaban

b) ¿Cuál es el número esperado de reprobación ?

3.- En un colegio determinado generalmente el 10 % de los estudiantes egresados no ingresa a la Universidad.

Hallar la probabilidad que de un total de 4 estudiantes egresados de ese colegio no ingresan a la Universidad.

- i) como mucho 3
- ii) entre 2 y 4
- iii) entre 1 y 3
- iv) 2 o más



- 4.- Si el 20% de los estudiantes tienen por lo menos un defecto físico. Determine la probabilidad de que de 4 estudiantes elegidos al azar :
- i) Uno tenga defecto física
  - ii) Ninguno tenga defecto físico
  - iii) A lo más dos tengan defecto físico
- 5.- Un estudiante tiene 2/5 de probabilidad de reprobar en un examen:
- a) Si realiza cuatro exámenes . Halle la probabilidad de que el estudiante repruebe en :
    - i) Dos exámenes
    - ii) Por lo menos en un examen
    - iii) Más de la mitad de los exámenes
  - b) Si en un mes realizó veinte exámenes  
¿ En cuántos cree usted que tendrá éxito ?

2.- Distribución normal

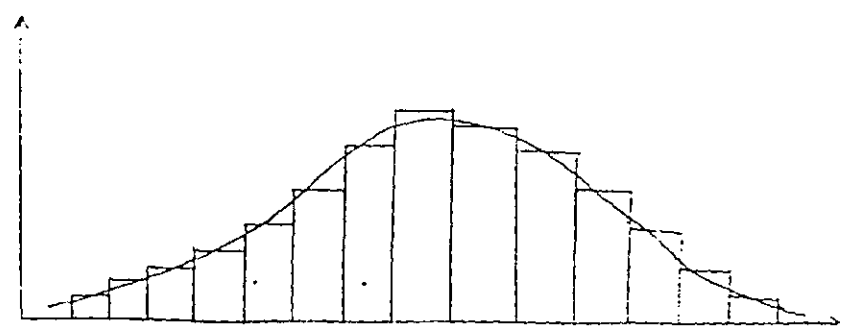
Una de las distribuciones de probabilidades continuas más importantes es la distribución normal o de Gauss Y viene dada por la función

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$$

donde  $\mu$  = media de la distribución  
 $\sigma$  = desviación estandar

Observaciones

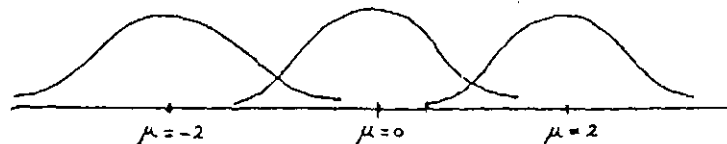
- i) La función  $f(x)$  de la distribución normal no se empleará para resolver problemas de aplicación
- ii) Hay una curva normal diferente para cada par de valores  $(\mu, \sigma)$  pero todos tienen las mismas características.
- iii) La distribución normal se utiliza cuando se tiene una variable en escala de intervalos continua con distribución de frecuencias simétrica y cuyo histograma tenga la forma de una campana



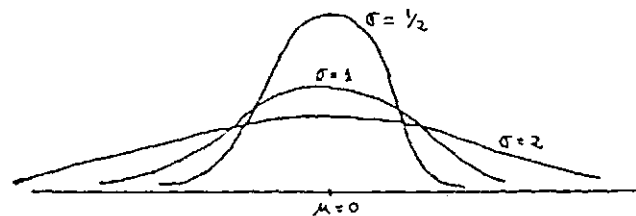
iv) La distribución queda completamente definida por el promedio y la desviación estandar. El promedio nos informa sobre la ubicación del centro de la distribución sobre el eje horizontal y la desviación estandar refleja la dispersión ( abertura de la campana )

Por ejemplo

si  $\sigma > 0$  fijo



si  $\mu = 1$  fijo



Definición

Si  $X$  es una variable que se distribuye normal con media  $\mu$  y desviación estandar  $\sigma$  entonces la variable estandarizada  $Z = \frac{X - \mu}{\sigma}$  se distribuye normal de media 0 y desviación estandar 1 y recibe el nombre de distribución normal estandar.

Cálculo de Probabilidades con la Distribución Normal Estandar

El área total limitada por la curva de la distribución normal estandar y el eje horizontal es uno :

El área bajo la curva comprendida entre  $-\infty$  y un valor específico  $Z$  lo denotaremos por  $\Phi(Z)$  y este valor aparece en la tabla del apéndice 1 para valores de  $Z$  entre  $-3$  y  $3$ .

Luego para el cálculo de probabilidades tenemos

- i)  $P(Z < a) = \Phi(a)$
- ii)  $P(a < Z < b) = \Phi(b) - \Phi(a)$
- iii)  $P(Z > b) = 1 - \Phi(b)$

Ejemplos

- 1.- Hallar  $\Phi(1,82)$  ;  $\Phi(1,96)$  ;  $\Phi(-2,58)$
- 2.- Calcular  $P(-0,47 < Z < 0,94) = \Phi(0,94) - \Phi(-0,47)$
- 3.- Calcular  $P(Z < 0,94) = \Phi(0,94)$
- 4.- Calcular  $P(1,96 < Z) = 1 - \Phi(1,96)$

Observación

- i) Si una variable  $X$  se distribuye normal de media  $\mu$  y varianza  $\sigma^2$

lo denotaremos por  $X \sim N(\mu, \sigma^2)$

ii) Cuando una variable  $X \sim N(\mu, \sigma^2)$  con  $\mu \neq 0$  y  $\sigma^2 \neq 1$  entonces para el cálculo de probabilidades es recomendable estandarizar la variable de la siguiente forma

$$\text{Si } X \sim N(\mu, \sigma^2) \text{ entonces } Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

luego :

$$P(X \leq a) = P\left(Z \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X \geq b) = P\left(Z \geq \frac{b - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

### Ejemplos

Si  $X \sim N(9,9)$  Hallar  $P(5 < X < 11)$  y  $P(X > 5)$

### Solución

$$\text{Si } X \sim N(9,9) \quad Z = \frac{X - 9}{3} \sim N(0,1)$$

$$P(5 < X < 11) = P\left(\frac{5 - 9}{3} < Z < \frac{11 - 9}{3}\right)$$

$$= \Phi(2/3) - \Phi(-4/3)$$

$$= 0,6536$$

$$P(X > 5) = P\left(Z > \frac{5 - 9}{3}\right)$$

$$= 1 - \Phi\left(-\frac{4}{3}\right)$$

$$= 1 - 0,0918$$

$$= 0,9982$$

### Ejercicios propuestos

1.- El peso verdadero de los estudiantes en una escuela tiene una distribución normal con media 32,8 kilos y desviación estandar 2,46 kilos ¿Cuál es la probabilidad de que un estudiante que está en esa escuela pese por lo menos 32,5 kilos ?

2.- Si  $Z \sim N(0,1)$ . Hallar

a)  $P(-1,2 < Z < 2,4)$

b)  $P(1,23 < Z < 1,87)$

c)  $P(-2,35 < Z < -0,5)$

d)  $P(Z \geq -1,64)$

e)  $P(-1,96 < Z < 1,96)$

3.- Si las alturas de 300 estudiantes se distribuye normalmente con media 68 pulgadas y varianza 9 pulgadas<sup>2</sup> ¿ Cuántos estudiantes tienen altura :

- i) Mayor de 72 pulgadas ?
- ii) Menor o igual a 64 pulgadas ?
- iii) Entre 65 y 71 pulgadas
- iv) Menor o igual a 68 pulgadas

4.- El Departamento encargado de evaluar a los estudiantes que ingresaron a un determinado colegio, los clasifica según su rendimiento en tres tipos A, B y C. El 15% de los estudiantes; los de mejor rendimiento son de la clase A, el 10%; los de peor rendimiento son de la clase C y el resto es de la clase B. Si suponemos que el rendimiento de los estudiantes se distribuye normalmente con una media de 30 puntos y una desviación estandar de 2,5 puntos y el Departamento reprueba inmediatamente a los alumnos de la clase C. Hallar :

- i) La puntuación mínima para que el estudiante sea clasificado como de la clase A.
- ii) La puntuación máxima para que el estudiante sea reprobado inmediatamente

5.- Supongamos que los puntajes de una prueba se distribuyen normalmente con una media de 22,5 puntos y una desviación estandar de 5,25 puntos. El profesor reprueba a todo aquel estudiante cuyo puntaje en la prueba es menor o igual a 20 puntos.

Hallar el porcentaje de reprobación en esa prueba.

### 3.- Distribución ji-cuadrado

Esta es una Distribución del tipo continua que se obtiene considerando muestras de tamaño "n" ; extraídas cada una de ellas de una población normal con desviación estandar .

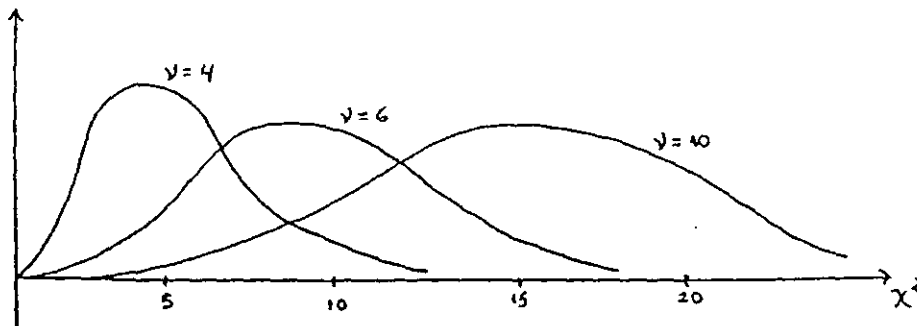
Si en cada muestra calculamos la expresión

$$\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\sigma^2}$$

tenemos que su distribución es una distribución muestral y se le conoce con el nombre de distribución ji-cuadrado y que se denota por  $\chi^2$

### Características

- i) Es una distribución asimétrica
- ii) Sólo toma valores positivos y es asintótica con respecto al eje horizontal ( $0 < \chi^2 < \infty$ )
- iii) Está caracterizada por un único parámetro  $\nu = n-1$  denominada "grados de libertad"
- iv) El área comprendida entre la curva y el eje horizontal es 1 (0 100%)
- v) Su gráfica para algunos valores de  $\nu$  es :

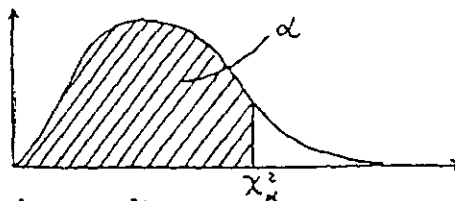


- vi) Si 
$$\chi^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\sigma^2}$$
 se distribuye como una ji-cuadrado

con  $\nu = n-1$  grados de libertad lo denotaremos por  $\chi^2 \sim \chi^2_{(n-1)}$

### Cálculo de percentiles Ji-cuadrado en tabla

Si el área comprendida entre 0 y  $\chi^2$  es  $\alpha$  entonces la tabla nos proporciona el valor del percentil  $\chi^2$  a un nivel  $\alpha$  con  $\nu$  grados de libertad.

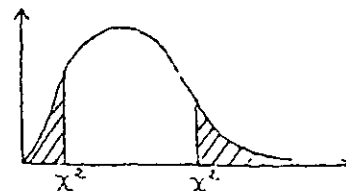


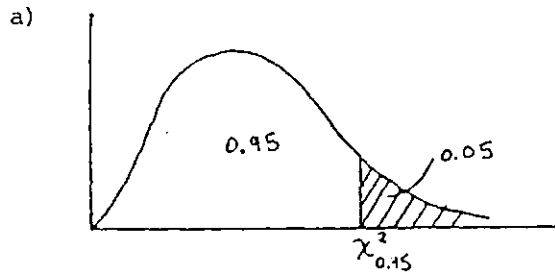
### Ejemplo resuelto

En la figura se muestra el gráfico de la distribución  $\chi^2$  con 5 grados de libertad.

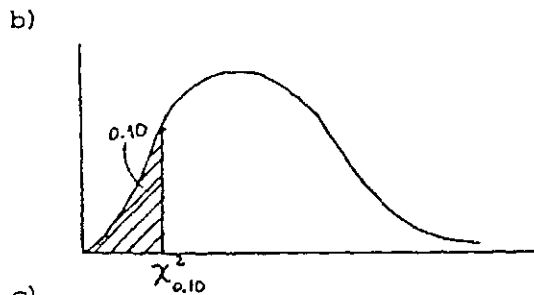
Hallar el valor del percentil correspondiente para el cual :

- a) El área sombreada de la derecha sea 0,05
- b) El área sombreada de la izquierda sea 0,20
- c) El área total sombreada sea 0.05

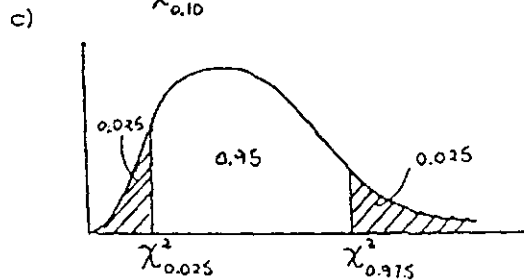


Solución

$$\chi^2_{0.05}(5) = 11.1$$



$$\chi^2_{0.10}(6) = 4.61$$



$$\chi^2_{0.025}(5) = 0.831$$

$$\chi^2_{0.975}(5) = 12.3$$

Ejercicios propuestos

1.- Para una distribución Ji-cuadrado con 12 grados de libertad hallar el valor del percentil  $\chi^2_p$  tal que

- a) El área de la derecha sea 0,05
- b) El área de la izquierda sea 0,99
- c) El área de la derecha sea 0,025

2.- Si la variable  $U \sim \chi^2_{(7)}$  y definimos  $P(U \leq \chi^2_p)$  = área comprendida entre 0 y el percentil  $\chi^2_p$

Hallar el valor de  $\chi^2_p$  tal que :

- a)  $P(U > \chi^2_p) = 0,025$
- b)  $P(U < \chi^2_p) = 0,05$
- c)  $P(\chi^2_{p1} \leq U \leq \chi^2_{p2}) = 0,90$

- 3.- Hallar la mediana de una distribución Ji-cuadra con : a) 9 g. l. b) 28 g. l. c) 40 g. l.

- 4.- Hallar los valores críticos de  $\chi^2_p$ , para los cuales el área en la cola derecha de la distribución Ji-cuadrado sea 0,05 si el N°de grado de libertad es : a) 8 b) 19 c) 40 d) 28

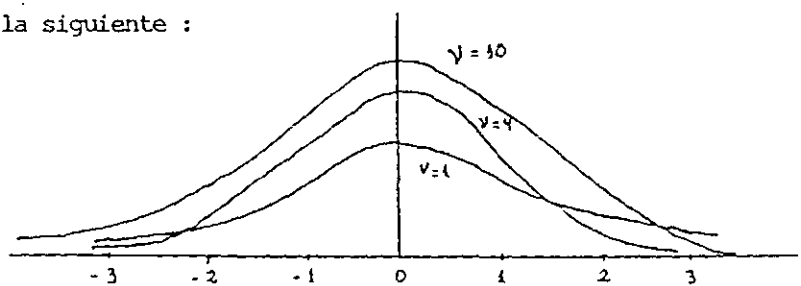
- 5.- Hacer el problema 4 si el área de la cola derecha es 0,01

4.- Distribución t de Student

Si se consideran muestras del tamaño n, extraídas de una población normal ( o aproximadamente normal ) con media  $\mu$  y si para cada muestra se calcula la expresión  $t^* = \frac{(\bar{X} - \mu)}{S} \sqrt{n - 1}$ ; utilizando la media muestral  $\bar{X}$  y la desviación estandar S; se obtiene una distribución muestral conocida con el nombre de " distribución t de Student " y diremos que  $t^*$  se distribuye como una t de student con  $\nu = n-1$  grados de libertad lo que denotaremos por  $t^* \sim t(\nu)$

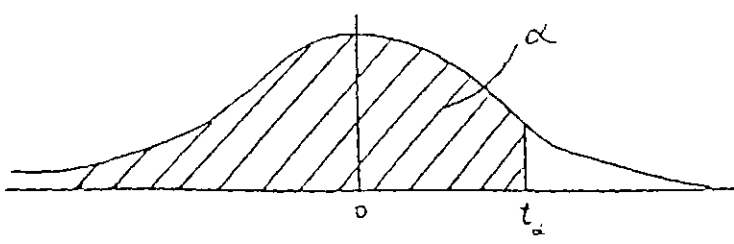
Observaciones

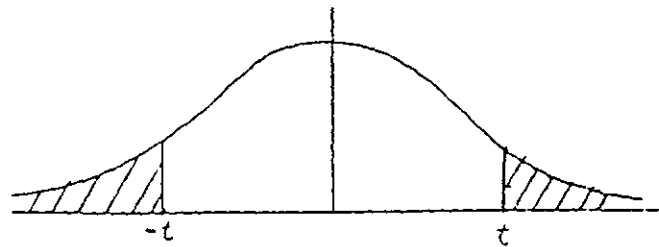
- i) Para grandes valores de  $\nu$  o n (  $n \geq 30$  ) la curva de la distribución t de student se aproxima a la curva de la distribución normal estandar
- ii) Para algunos valores de  $\nu$  la curva de la distribución t de student es la siguiente :



Cálculo de percentiles  $t_\alpha$  mediante tablas

El área total comprendida bajo la curva y el eje horizontal es uno. Si el área comprendida entre  $-\infty$  y  $t_\alpha$  es  $\alpha$  entonces la tabla nos da el valor del percentil  $t_\alpha$  a un nivel  $\alpha$  con  $\nu$  grados de libertad.





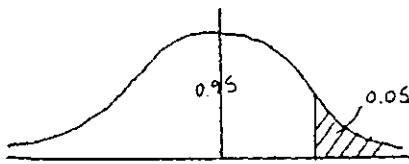
Ejemplo resuelto

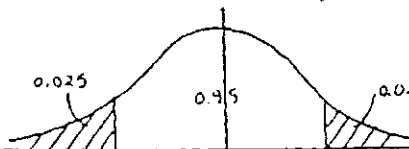
En la figura se muestra el gráfico de una distribución t de student con 9 grados de libertad.

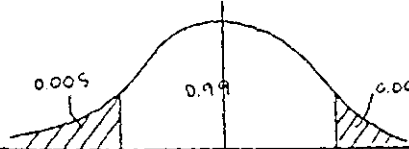
Hallar el valor t para el cual :

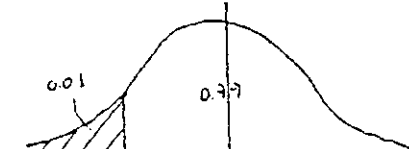
- a) El área sombreada de la derecha sea 0,05
- b) El área total sombreada sea 0,05
- c) El área total no sombreada sea 0,99
- d) El área sombreada de la izquierda sea 0,01
- e) El área de la izquierda de t sea 0,90

Solución

a)   $t = t_{0.05}^{(9)} = 1.83$

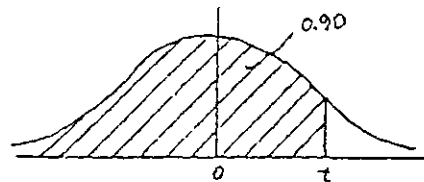
b)   $t = t_{0.025}^{(9)} = 2.26$

c)   $t = t_{0.005}^{(9)} = 3.25$

d)   $t = t_{0.01}^{(9)} = 2.32$



e)



$$t = t_{0.90}^{(4)} = 1.38$$

### Ejercicios propuestos

- 1.- Hallar los valores críticos de t para los cuales el área de la cola derecha de la distribución t de student sea 0,05 si el número de grados de libertad es igual a :
  - a) 16    b) 27    c) 5
  
- 2.- Hallar los valores de t para la distribución t de student con grados de libertad que satisface cada una de las condiciones siguientes :
  - a) El área entre -t y t sea 0,90 y  $n = 25$
  - b) El área a la izquierda de -t sea 0,025 y  $n = 20$
  - c) El área a la derecha de t más el área a la izquierda de -t sea 0,01 y  $n = 5$
  - d) El área a la derecha de t sea 0,55 y  $n = 16$
  
- 3.- Para una distribución t de student con 15 grado de libertad hallar el valor t tal que :
  - a) El área a la derecha sea 0,01
  - b) El área a la derecha sea 0,10
  - c) El área a la izquierda sea 0,95
  - d) El área conjunta a la derecha de t y a la izquierda de -t sea 0,01

## VIII MUESTREO PROBABILISTICO

Definición de la población y creación del marco muestral.

- . Posición de esta etapa entre las etapas principales:

DEFINICION DE LOS OBJETIVOS

DEFINICION DE LA POBLACION Y  
Y CREACION DEL MARCO MUESTRAL

CONCEPCION DE LA BOLETA

SELECCION DE LA MUESTRA

RECOLECCION DE LOS DATOS

PROCESAMIENTO DE LOS DATOS

ESTIMACION

- . Importancia de definir bien los objetivos

- . Tres tipos de objetivos:

- . Objetivos ideales::: las metas originales de la encuesta
- . Objetivos definidos: los que parecen razonable desde el punto de vista de las operaciones.
- . Objetivos actuales : los que son finalmente los resultados que se propone obtener.

Plan de muestreo

- . DISEÑO MUESTRAL
- . METODO DE ESTIMACION
- . METODOS DE EVALUACION DEL ERROR MUESTRAL

Elementos del diseño muestral (para la extracción de la muestra)

- . Población objeto y población sujeto
- . Marco muestral
- . Unidades de muestreo
  - . Unidad de análisis
    - . Unidad de respuesta
      - . Unidad de referencia
- . Tamaño de la muestra
- . Métodos de selección de la muestra

### Métodos de estimación (para la inferencia a la población)

- . Como calcular las estimaciones de las características poblacionales a partir de la muestra.

### Evaluación del error muestral (medición de la confiabilidad)

- . Procedimientos para estimar los errores muestrales, particularmente la varianza muestral.

### Factores que se deben considerar en relación con el plan de muestreo

- . Consideraciones operacionales y administrativas
- . Calidad de los datos disponibles.
- . Confiabilidad requerida.
- . Costos

### Definición de la población

- . El conjunto de todas las unidades que se quiere estudiar
- . A menudo se llama el universo de la encuesta
- . Depende de los objetivos de la encuesta (y vice versa)
- . Se necesitan definiciones claras:
  - . de las características de las unidades.
  - . de la ubicación geográfica de las unidades
  - . del período de referencia de la encuesta
- . A menudo una buena definición de la población es muy difícil. Casi siempre se puede describir en términos generales, pero es muy difícil determinar si ciertas unidades pertenecen o no a la población.

### Población objeto

- . La población que se quisiera estudiar
- . El Conjunto de unidades al cual los resultados de la encuesta deberían aplicarse.

### Población sujeto (muestreada)

- . La población realmente cubierta por la encuesta
- . La población de cuyas unidades se puede recolectar datos, teniendo en cuenta las restricciones operacionales
- . Entonces el conjunto de unidades a lo cual se aplican realmente los resultados de la encuesta
- . Frecuentemente bastante distinto de la población objeto
- . Sucede a menudo que el usuario de los resultados no entiende la diferencia entre estas dos poblaciones.
- . Importancia de bien documentar las diferencias

### Características deseables de la población

- . Unidades claramente definidas
- . Población conforme con los objetivos de la encuesta
- . Población sujeto lo más cerca posible de la población objeto
- . Período de referencia bien especificada
- . Disponibilidad de datos necesarios para la definición del marco muestral

### Marco Muestral

- . Es la lista de todos los elementos de la población sujeto
- . Es la herramienta concreta que permite la identificación de las Unidades que se deben estudiar
- . Determina la manera de acceder a la población
- . Es a partir del marco muestral que se seleccionan las unidades a estudiar

### Características deseables del marco muestral

- . Debería contener la lista de todos los elementos de la población sujeto (en el caso de un censo como en el caso de una encuesta por muestreo)
- . Debería identificar cada elemento de manera precisa
  - . ubicación municipal, número de teléfono, nombre y apellidos, ..., según el caso
- . Puede incluir información adicional
  - . método preferido de contacto
  - . información relevante a la estratificación
  - . información sobre el tamaño o la importancia de la unidad
  - . referencia a otros listados o archivos
  - . la fecha de la última verificación de los demás datos

El marco muestral tiene un papel importante en:

- . La definición de la población sujeto
- . El método de recolección de datos
- . El método de selección de la muestra

### Creación del marco muestral

- . La población sujeto está dividida en unidades de muestreo
- . El marco muestral es una lista de todas las unidades de muestreo, con algunas de sus características

. Hay tres casos:

- . Caso simple: La unidad de muestreo es la unidad (el elemento) de la población.

Ejemplo : Población sujeto: los alumnos de una universidad  
 Unidad de la población : un alumno  
 Unidad de muestreo : un alumno  
 Marco muestral: lista de todos los alumnos

. Caso medio: La unidad de muestreo es un grupo de unidades de la población

Ejemplo : Población sujeto: los residentes de una ciudad, de 18 o más años de edad.  
 Unidad de la población: persona de 18 o más años de edad, residente en la ciudad  
 Unidad de muestreo: vivienda (u hogar)  
 Marco muestral: lista de todas las viviendas de la ciudad (o lista de todos los hogares)

Caso complejo: Marco muestral multi-etápico, teniendo cada etapa su unidad de muestreo

Ejemplo: Población sujeto: los residentes de una ciudad  
 Unidad de la población: residente de la ciudad  
 No hay listado de los residentes, ni de las viviendas

Tres niveles de unidades de muestreo:

- . sector de empadronamiento censal
- . manzana
- . vivienda

Las etapas:

- . lista de sectores
- . selección de algunos sectores
- . lista de manzanas en cada sector  
seleccionado
- . selección de manzanas
- . lista de viviendas en cada manzana  
seleccionada
- . selección de viviendas

Se trata realmente de una jerarquía de marcos parciales

Se pueden distinguir cuatro tipos de marco muestral:

Marco en forma de lista

- . Se prepara una lista de todas las unidades de muestreo

- . Normalmente corresponde a los casos "simple" y "medio" citados arriba

#### Marco de áreas

- . Lista de áreas de enumeración, es decir una lista de unidades geográficas
- . Normalmente implica más de un nivel de unidades
- . Corresponde al caso "complejo" de unidades muestrales
- . Importancia de la cartografía

#### Otro

- . Por ejemplo, permutaciones de cifras para la selección aleatoria de número de teléfono
- . La relación entre las unidades del "marco" y las unidades de la población puede ser muy indirecta

#### Combinado

- . Jerárquico
  - . Ejemplo: Primera etapa: es un marco de áreas, digamos de sectores censales. Se seleccionan varios sectores.  
Segunda etapa: una lista completa de las viviendas en cada sector seleccionado
- . Múltiple
  - . A menudo, es posible utilizar una lista de una parte importante de la población, pero la lista queda incompleta.  
Se puede utilizar un marco de áreas para completar la cobertura de la población
  - . Ejemplo: Encuesta sobre la producción de vino en Chile.  
Hay una lista de los 25 productores más importantes, representando quizás 85 - 90% de la producción nacional.  
Para cubrir los pequeños productores, se puede utilizar también un marco de áreas basado en las partes rurales del país.

#### Factores que afectan la selección del marco muestral

- . Los objetivos de la encuesta
- . La población sujeto
- . Los costos
- . Datos administrativos disponibles
- . Otros datos externos disponibles

- . Recursos humanos
- . Recursos materiales
- . Calendario

Es muy difícil contar con un marco muestral que corresponda bien a la población sujeto

#### Problemas de los marcos muestrales

Dos problemas mayores:

- . Mantener al día la cobertura del marco
- . Mantener al día la clasificación de las unidades

Los tres errores posibles en cada caso:

- . Elementos que faltan (sub-cobertura, omisión)
  - . Nacimientos, inmigraciones, divisiones
- . Elementos ajenos (sobre-cobertura)
  - . Muertes, emigración, fusiones
- . Repeticiones
  - . Elementos correctos, incluidos dos o más veces en el marco

#### ¿Cómo poner al día la cobertura?

Cuatro posibilidades:

- . La encuesta misma
  - . Ejemplo: Productores de vino en quiebra, descubiertos al momento de hacer las entrevistas sobre la producción
- . Esfuerzo especial para poner al día los sectores que se van a seleccionar para una encuesta en el futuro
  - . Personal del instituto de estadística visita algunas regiones y prepara listas de todas las viviendas en ciertos sectores
- . Nuevas listas administrativas
  - . Ejemplo: lista anual de todas las clínicas, por provincia y municipio, proporcionada por el ministerio de salud
  - . Nuevos datos administrativos
    - . Ejemplo: ministerio de hacienda proporciona una lista de todas las empresas realizando exportaciones de más de \$ 1.000.000 en el año 1986

#### ¿Cómo reconocer unidades distintas?

- . Ejemplo: en el listado de restaurantes de Santiago, se encuentran:
 

. Pisa 4 Vientos	y	Pizza Cuatro Vientos
Apumarque		Apoquindo 6529

- . ¿Cómo se puede saber que se trata de la misma unidad?
- . Tal listado puede incluir 15.000 restaurantes, por lo tanto se necesita algo más metódico que la opinión de un empleado que conoce varios restaurantes en su barrio, a fin de evitar problemas graves de omisión o repetición

#### Clasificación de las unidades

- . Importancia de las definiciones y de los sistemas de clasificación
- . Mala clasificación de una unidad puede dar omisión o sobre-cobertura
  - . Ejemplo: En una encuesta de hogares, cuya última unidad de selección es la vivienda privada ocupada como un lugar comercial (por que hay un bar en el primer piso) es un caso de omisión; la clasificación de una tienda no ligada a una vivienda privada, como vivienda, sería un caso de sobre-cobertura

#### ¿Cómo poner al día la clasificación?

##### Tres posibilidades

- . La encuesta misma
  - . Durante la recolección de los datos, se pueden incluir algunas preguntas para verificar la clasificación de la unidad, y eliminar las que no corresponden. Así se puede reducir la sobre-cobertura
- . Otras encuestas
  - . Si todas las encuestas incluyen preguntas sobre la clasificación, las unidades de tipo "B" encontradas en la encuesta "A" se pueden entregar a la encuesta "B" y vice versa
  - . Sin embargo, este método tiene grandes problemas de coordinación de práctica
- . Datos administrativos
  - . A veces incluyen información que se puede utilizar para verificar la clasificación en el marco muestral

#### Unidades de encuesta

- . Unidad de muestreo :           unidad seleccionada
- . Unidad de análisis :           unidad de descripción o inferencia para la población
- . Unidad de respuesta:           unidad que proporciona las respuestas
- . Unidad de referencia:        unidad a la que se refieren los datos recolectados



Ejemplo:

- . Objetivo: Encuesta sobre la mano de obra (población económicamente activa)
- . Unidad de muestreo : vivienda
- . Unidad de análisis : individuos y familias
- . Unidad de respuesta: una persona responsable que se encuentra en la vivienda
- . Unidad de referencia: cada miembro del hogar, de más de 14 años de edad

Ejemplo:

- . Objetivo: Recolectar datos sobre los hogares (familias extendidas)
- . Unidad de muestreo : vivienda
- . Unidad de análisis : hogar (familia extendida)
- . Unidad de respuesta: jefe de hogar, o de familia nuclear
- . Unidad de referencia: familia nuclear

#### Muestreo probabilístico

- . Cada unidad de la población tiene una probabilidad no-nula de ser seleccionada
- . Esta probabilidad se puede medir
- . La muestra se selecciona de manera aleatoria
- . Se utilizan las probabilidades de selección para calcular las estimaciones para toda la población a partir de la muestra

Hay varios métodos de muestreo probabilístico. Un diseño muestral puede utilizar uno de ellos, o una combinación de varios.

#### Ventajas del muestreo probabilístico

- . Objetividad
- . Los métodos de estimación tienen una justificación teórica
- . El método de muestreo determina el método de estimación
- . El error muestral se puede medir
- . Normalmente, el error muestral es más pequeño que en el caso de muestreo no-probabilístico

### Muestreo aleatorio simple (MAS)

- . Una muestra con tamaño pre-determinado se selecciona de una lista de todas las unidades de la población
- . Cada unidad tiene la misma probabilidad de selección
- . Si se selecciona una sola unidad de una población de  $N$  unidades, cada una tiene una probabilidad de  $1/N$  de ser seleccionada
- . Para una muestra de más de una unidad, cada unidad tiene la misma probabilidad de selección, y cada combinación de unidades tiene la misma probabilidad de selección
- . Entonces cada muestra de  $n$  unidades tiene la misma probabilidad de selección, y cada unidad tiene una probabilidad de selección de  $n/N$ .
- . Se saca una unidad a la vez hasta que se tenga la muestra de tamaño deseado

Hay dos métodos de muestreo aleatorio simple:

#### Con reposición (MASCR)

- . Cada unidad de la población queda disponible para cada selección
- . La misma unidad puede ser incluida varias veces en la muestra

#### Sin reposición (MASSR)

- . Una vez seleccionada, una unidad no se puede seleccionar otra vez en la misma muestra, entonces la muestra tiene unidades distintas

Este método es lo más utilizado en la práctica

#### Ventajas del MAS

- . Su simplicidad
- . MASSR da estimaciones más precisas que MASCR

#### Desventajas del MAS

- . MASCR permite duplicación
- . Hay que seleccionar muchos números aleatorios
- . MASCR ni MASSR hacen uso de información adicional disponible sobre la población

#### Información necesaria para el marco muestral

- . Sólo una lista que identifique cada unidad de la población

## Ejemplo del MAS

Población : Un pueblito donde viven seis hombres:

Arturo, Bernardo, Carlos, David, Edmundo y Francisco

Muestra : MASSR de dos hombres para hacer un trabajo

Hay solamente 15 muestras posibles:

AB	BC	CD	DE	EF
AC	BD	CE	DF	
AD	BE	CF		
AE	BF			
AF				

- . Cada persona tiene una probabilidad de selección de  $5/15 = 1/3$
- . Cada pareja (= cada muestra) tiene probabilidad de selección de  $1/15$
- . Como sacar la muestra: Dar a cada persona un número, por ejemplo
 

A	B	C	D	E	F	se hace	1	2	3	4	5	6
---	---	---	---	---	---	---------	---	---	---	---	---	---
- . Utilizar una tabla de números aleatorios
- . Empezando al comienzo de la primera fila de la tabla, se encuentran las cifras 5939 ...; entonces Edmundo (5) y Carlos (3) harán la tarea
- . Si se hubiera empezado con la fila 20, columna 55, la muestra sería Bernardo y Arturo (número aleatorios 8227010 ...); si fuera una selección por MASSCR, el pobre Bernardo tendría que hacer ambas mitades del trabajo.

## TABLA DE NÚMEROS ALEATORIOS

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
00	59391	58030	52098	82718	87024	82848	04190	96574	90464	29065
01	99567	76364	77204	04615	27062	96621	43918	01896	83991	51141
02	10363	97518	51400	25670	98342	61891	27101	37855	06235	33316
03	06059	19558	64432	16706	99612	59798	32803	67708	15297	26612
04	11258	24591	36863	55368	31721	94335	34936	02566	00972	08188
05	95068	88628	35911	14530	33020	80428	39936	31855	34334	64865
06	54463	47237	73800	91017	36239	71824	83671	39892	60518	37092
07	16874	62677	57412	13215	31389	62233	80827	73917	82002	84420
08	92494	63157	76593	91316	03505	72389	96363	52887	01087	66091
09	15669	56689	35682	40844	53256	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	86211	36584	67466	69373	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	80327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
25	83266	32883	42451	15579	38155	29793	40914	65990	16255	17777
26	76970	80876	10237	39515	79152	74798	39357	09054	73579	92359
27	37074	65190	44785	68624	98336	84481	97610	78735	46703	98265
28	83712	06514	30101	78295	54656	85417	43189	60048	72781	72606
29	20287	56862	69727	94443	64936	08366	27227	05158	50326	59566
30	74261	32592	86538	27041	65172	85532	07571	80609	39285	65340
31	64081	49863	08478	96001	18888	14810	70545	89755	59064	07210
32	05617	75818	47750	67814	29575	10526	66192	44464	27058	40467
33	26793	74951	95466	74307	13330	42664	85515	20632	05497	33625
34	65988	72850	48737	54719	52856	01596	03845	35067	03134	70322
35	27366	42271	44300	73399	21105	03280	73457	43093	85122	48657
36	56760	10909	98147	34736	33863	95256	12731	66598	50771	83665
37	72880	43338	93643	58904	59543	23943	11231	83268	65938	81581
38	77888	38100	03062	58103	47961	83841	25878	23746	55903	44115
39	28440	07819	21580	51459	47971	29882	13990	29226	23608	15873

## Eficiencia

- Para comparar los distintos métodos de muestreo, se usa el concepto de eficiencia o eficiencia relativa
- Si la varianza muestral del método "A" es más pequeño que la del método "B", con muestra del mismo tamaño, se dice que "A" es más eficiente que "B"
- La eficiencia relativa de "A" (con respecto a "B") es simplemente:

$$\frac{\text{varianza según "B"}}{\text{varianza según "A"}}$$

- Si se dice que el método "A" es más eficiente que el método "B", quiere decir que las estimaciones de "A" son más precisas que las de "B"

La identificación de "eficiencia" y "precisión" en el párrafo anterior corresponde al uso común y corriente en la estadística, cuando normalmente se consideran sólo estimadores insesgados. Sin embargo, en el caso general, se trata de dos casos distintos

Hay que distinguir dos términos:

**Precisión :** Se trata de considerar la variabilidad de las observaciones con referencia al promedio de las mismas observaciones, tomando en cuenta todas las muestras posibles.

La medida básica de precisión es la varianza (V)

**Exactitud :** Se considera la variabilidad de las observaciones con referencia al valor verdadero en la población, tomando en cuenta todas las muestras posibles

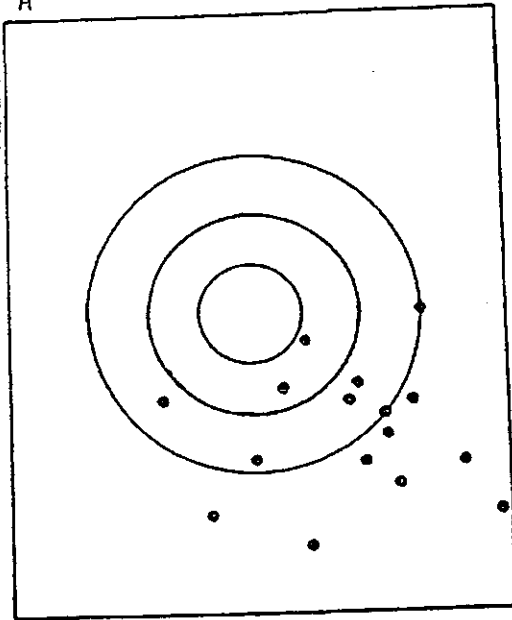
La medida básica de exactitud es la media cuadrática de error (MCE)

Si un estimador es insesgado (tiene sesgo  $B = 0$ ), su precisión y exactitud son lo mismo, dado que  $MCE = V + B^2$

Estrictamente, la idea de eficiencia se debería identificar con la exactitud, entonces digamos que:

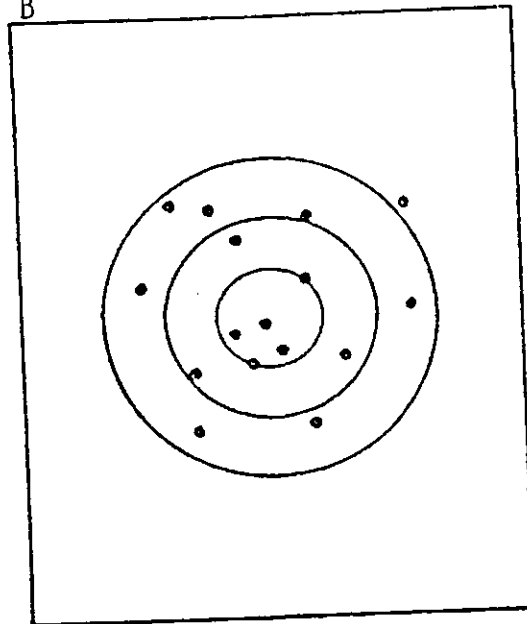
Entre dos estimadores, el estimador con menor varianza es más preciso, y el estimador con menor MCE es más eficiente, o tiene mejor exactitud

A



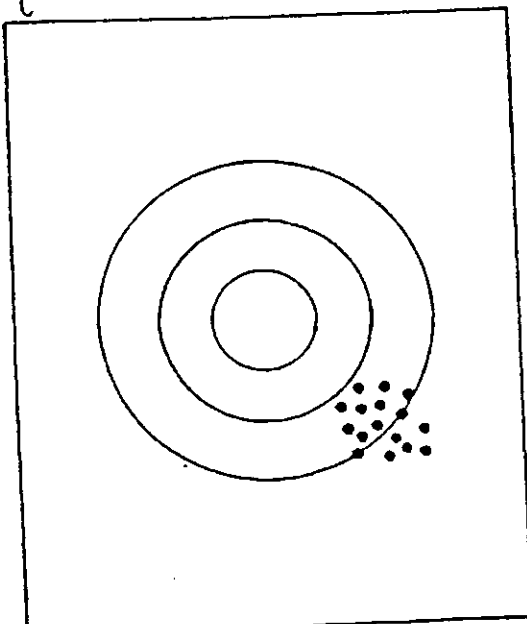
BAJA PRECISION  
BAJA EXACTITUD  
(SESGO GRANDE)

B



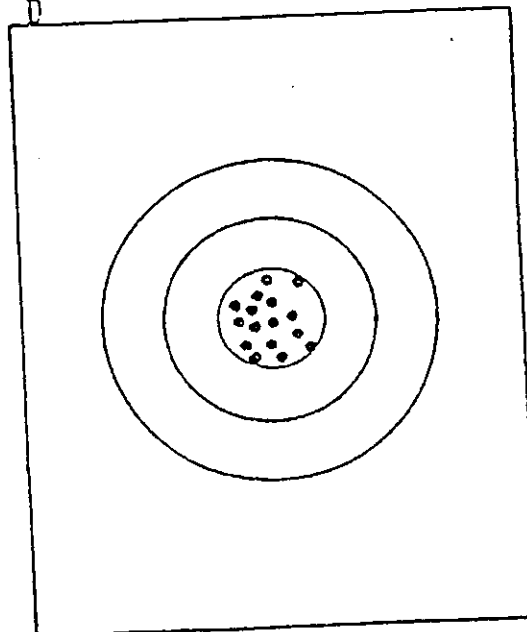
BAJA PRECISION  
BAJA EXACTITUD  
(SESGO PEQUEÑO)

C



ALTA PRECISION  
BAJA EXACTITUD  
(SESGO GRANDE)

D



ALTA PRECISION  
ALTA EXACTITUD  
(SESGO PEQUEÑO)

Factores que determinan precisión y exactitud

- . Error muestral
- . Error no-muestral, o error de observación

Error muestral

- . Error atribuible al hecho de observar sólo una muestra (una fracción) de la población
- . Fuentes posibles:
  - . Tamaño de la muestra y de la población
  - . Variabilidad de las características estudiadas en la población
  - . Diseño muestral
  - . Método de estimación
- . Salvo en el caso de un verdadero censo, siempre hay varianza muestral
- . También puede haber sesgo muestral

Error no muestral

- . Existe en los censos como en las encuestas por muestreo
- . También puede incluir componentes de varianza y de sesgo
- . Fuentes posibles:
  - . Errores de cobertura del marco muestral
  - . Errores de clasificación
  - . Errores de respuesta y de no-respuesta
  - . Imprecisión de las preguntas, problemas de interpretación
  - . Sesgo de los entrevistadores
  - . Mal período de referencia
  - . Mala aplicación de los procesos de recolección de los datos
  - . Errores de transcripción o de codificación
  - . Errores de digitación
  - . Errores de control y verificación
  - . Errores de imputación

### Muestreo sistemático

- . Necesita un intervalo de muestreo y un número de arranque aleatorio
- . El intervalo de muestreo es  $K = n/N$
- . El arranque aleatorio es un número entero  $R$  entre 1 y  $k$
- . Las unidades  $R, R+k, R+2k, \dots, R+(n-1)k$  son seleccionadas

Ejemplo:  $N = 153.000$  granjas  
 $n = 9.000$  "  
 $k = 153.000 / 9.000 = 17$

Si se selecciona el arranque aleatorio  $R = 4$ , las granjas con número 4, 21, 38, 55, 72, 89, 104, ... , 152.987 son seleccionadas

- .  $R$  tiene la probabilidad  $1/k = 1/17$  de ser cada uno de los valores 1, 2, ... ,  $k$
- . Hay  $k$  muestras posibles, cada una con probabilidad de selección  $1/k$
- . La probabilidad de selección de cada unidad de la población es  $1/k = n/N$ .
- . Esto es el muestreo sistemático lineal
- . Generalmente,  $N$  no es múltiple exacto de  $n$
- . Sin embargo, hay varias modificaciones sencillas
- . El tamaño del intervalo muestral puede ser redondeado, tomando el valor entero inferior o superior a  $k$
- . Esto puede dar una muestra un poco más pequeño o más grande que  $n$ , según el valor de  $R$ .
- . Se puede utilizar el muestreo sistemático circular, con  $R$  seleccionado aleatoriamente entre 1 y  $N$
- . Ejemplo .  $N = 11, n = 3, k = 4$  (redondeado)
  - . Supongamos que  $R = 10$
  - . Las unidades seleccionadas tendrían las posiciones 10, 14, 18 pero la población tiene sólo 11 unidades
  - . Las unidades realmente seleccionadas son 10, 14-11, y 18-11, es decir 3, 7 y 10
- . Para mejorar la cobertura muestral de la lista, en la práctica no se redondea  $k$  directamente, sino los valores  $R, R+k, R+2k, \dots$ , individualmente

### Ventajas del muestreo sistemático

- . La selección es más simple que con MAS, porque hay que sacar sólo un número aleatorio
- . A menudo proporciona estimaciones más precisas que MAS, siendo la muestra distribuida de manera más uniforme en la población



#### Ventajas del muestreo sistemático

- . La selección es más simple que con MAS, porque hay que sacar sólo un número aleatorio
- . A menudo proporciona estimaciones más precisas que MAS, siendo la muestra distribuida de manera más uniforme en la población

#### Desventajas

- . Periodicidad : Si la lista está ordenada de tal manera que la característica por medir tiene un patrón cíclico cuyo período está en relación con el intervalo muestral, la precisión de la estimación puede ser muy mala
- . No hay estimador insesgado de la varianza muestral

#### Información necesaria para el marco muestral

- . Una lista de todas las unidades de la población

### Muestreo estratificado

- . Utiliza información disponible para mejorar la representación de los varios grupos de la población
- . Si es conocido (o sospechado) que la variable de interés está correlacionada con otra variable, y si la información sobre la segunda variable está disponible para cada unidad muestral, la población se puede dividir en grupos que son los más homogéneos posibles, respecto a esta variable
- . Estos grupos se llaman estratos
- . Cada estrato se trata como una sub-población independiente y una parte de la muestra se selecciona independientemente dentro de cada estrato
- . El método de selección de la muestra puede ser diferente entre los distintos estratos
- . El muestreo estratificado proporciona no solamente estimaciones para toda la población, sino que también separadamente para cada estrato o grupo de estratos
- . Las variables usadas para la estratificación pueden ser de naturaleza geográfica o no geográfica

### ¿Porque estratificar? (Ventajas)

- . Para mejorar la precisión de las estimaciones para toda la población
- . Para obtener estimaciones para cada estrato
- . Por razones administrativas
- . Para aprovechar diferentes métodos de muestreo en varias partes de la población

### Factores que determinan la eficiencia de la estratificación

- . La selección de la (s) variable (s) de la estratificación
- . El número de estratos
- . Los límites de los estratos
- . La afijación de la muestra entre los estratos

### Desventajas

- . Más complejo que MAS o muestreo sistemático puro
- . Información para estratificación debe estar disponible para cada unidad, y debe ser más o menos precisa

### Información necesaria para el marco muestral

- . Lista de todas las unidades de la población

- y
- . Valor(es) de variable(s) de estratificación para cada unidad
- c
- . Posibilidad de una división de la población en unidades geográficas

#### Afijación (alocación) de la muestra entre los estratos

- . Afijación proporcional
  - . La misma fracción de muestreo en cada estrato
  - . El tamaño de la muestra es proporcional al número de unidades en el estrato
- . Afijación no proporcional
  - . Si se sabe que la variabilidad de la(s) variable(s) de interés es muy diferente entre los varios estratos, o si los costos de la recolección de los datos son muy diferentes, estos factores se pueden tomar en cuenta
  - . Hay varios métodos de afijación, por ejemplo:
    - . Afijación óptima: Método complejo que aumenta el tamaño de la muestra en estratos con mayor varianza o menor costo
    - . Afijación de Neyman:
      - Como la fijación óptima pero utiliza sólo información sobre la varianza
    - . Afijación ... : (sin nombre especial pero quizás lo más útil en la práctica): como la afijación óptima pero utiliza sólo información sobre los costos
- . Afijación proporcional a "X"
  - . Como la afijación proporcional simple pero utiliza información sobre otra variable correlacionada con la variable de interés, a menudo información sobre el tamaño de las unidades

### Muestreo por conglomerados

- . Requiere una población que puede ser estructurada de manera jerárquica
  - . Ejemplo :      Persona viven en viviendas  
                   Un grupo de viviendas es una manzana  
                   El conjunto de manzanas es una ciudad
- . Un conglomerado es una unidad de muestreo que incluye varias unidades de la población
  - . Ejemplo :      Una vivienda se puede ver como un conglomerado que consta de varias personas
    - o Una manzana se puede ver como un conglomerado de varias viviendas
- . El muestreo por conglomerados es un método que selecciona conglomerados enteros como las últimas unidades muestrales, e incluye en la muestra todas las unidades de la población que se encuentran dentro de los conglomerados seleccionados
- . La selección de los conglomerados se puede hacer por MAS, muestreo sistemático o cualquier otro método probabilístico, si la información necesario para la aplicación del método está disponible

### Ejemplos de poblaciones con conglomerados

Población estudiada	Variable de interés	Unidad de la población	Conglomerado
Escuela secundaria	Planes de Carrera	Alumno	Clase
Tráfico por un puente	Origen y destino	Auto	Período de tiempo
Aeropuerto	Gastos de los turistas	Pasajeros saliendo del país	Vuelo (avión)

### Ventajas

- . Se puede utilizar aún cuando no hay lista de las unidades de la población
- . No es necesario preparar una lista completa
- . Si la recolección de los datos se hace por entrevistas personales, es posible reducir los costos y el tiempo de viaje de manera importante, especialmente en el caso de poblaciones rurales muy dispersas

## Desventajas

- Menor precisión, generalmente, debido a la tendencia de las unidades vecinas a ser más o menos semejantes con respecto a sus características de interés

## Información necesaria para el marco muestral

- Una lista de los conglomerados
  - o
- La posibilidad de construirla
- El valor del muestreo por conglomerados, sobre todo desde el punto de vista de la precisión, depende mucho de la relación entre la naturaleza del conglomerado y la variable estudiada
  - Ejemplo : Los conglomerados son hogares, siendo la unidad de la población la persona
    - La variable de interés puede ser el nivel material de vida, o la religión de los individuos
    - Se puede suponer que no vale la pena entrevistar a cada miembro del hogar, porque la respuesta de cada uno sería la misma
    - Es decir que se puede esperar una correlación intraclase muy fuerte
    - Entonces todo el conglomerado no puede proporcionar en realidad mucho más información que cualquier individuo del grupo hablando para todos
    - Si se entrevista a todos los individuos, el tamaño efectivo de la muestra (en términos de la información obtenida) es muy reducido, comparando a una muestra aleatoria del mismo número de personas
    - Entonces este diseño tendría una varianza elevada, comparada con la varianza de la muestra aleatoria simple
    - Se dice que el diseño tiene un alto "efecto del diseño", es decir que tiene baja eficiencia
  - Al revés, si la correlación entre los elementos del conglomerado es negativa, como sería el caso en un estudio de la distribución de la población por sexo y edad, un conglomerado puede efectivamente proporcionar más información que un número equivalente de individuos seleccionado independientemente
    - En este caso el efecto del diseño es menor que uno, es decir que el diseño por conglomerados es más eficiente que un diseño aleatorio simple

- . Entonces, desde el punto de vista de la varianza, un diseño muestral basado en conglomerados es más eficiente cuando cada conglomerado es lo más heterogeneo posible, o sea cuando cada conglomerado representa más o menos una "mini-población"
- . Comparar con el caso de estratificación, donde el diseño es más eficiente cuando los estratos son homogéneos, y la mayor parte de la varianza se encuentra entre los estratos
- . Cada estrato proporciona una parte de la muestra; si el estrato es información sobre su contenido
- . En el caso de los conglomerados, sólo algunos son seleccionadas, y todas sus unidades deben ser entrevistadas, entonces es preferible que cada conglomerado de la mayor información posible, es decir que el conglomerado sea heterogeneo, y que hayan diferencias entre los distintos conglomerados
- . Si se sospecha homogeneidad dentro de los conglomerados (es decir correlación intraclase positiva), es preferible buscar un número más grande de conglomerados pequeños

## Muestreo multi-etápico

- . Indica cualquier diseño utilizando dos o más niveles o etapas de muestreo
- . Requiere una jerarquía de dos o más niveles de unidades de muestreo
- . El proceso general:
  - . Construir una lista completa de unidades de la primera etapa (unidades primarias)
  - . Utilizando un método de muestreo apropiado, seleccionar una muestra de unidades primarias
  - . Dentro de cada unidad primaria seleccionada, y sólo dentro de estas unidades, construir una lista de todas las unidades de la segunda etapa (unidades secundarias)
  - . Seleccionar una muestra de unidades secundarias dentro de cada una de las unidades primarias seleccionadas
  - . Si hay más de dos etapas, dentro de cada unidad secundaria seleccionada, hacer una lista de todas las unidades del tercer nivel
  - . Seleccionar una muestra de ellas dentro de cada unidad secundaria ya seleccionada, etc.
  - . La etapa final puede ser la segunda, tercera, cuarta, ... según el diseño
- . En casos reales, diseños con dos o tres etapas son comunes; es muy raro encontrar un diseño con más de cuatro etapas
- . A menudo, la última unidad muestral es el elemento de la población
- . Sin embargo, en la mayoría de las encuestas de hogares, la última unidad muestral es el hogar o la vivienda mientras el elemento de la población es la persona
- . Un buen marco muestral es necesario para cada etapa
- . Generalmente, el marco de unidades primarias es más completo, y más establece en el tiempo, que los marcos de unidades de las últimas etapas, es decir las unidades más pequeñas
  - . Ejemplo: Si las etapas de un diseño son provincia: sector censal: hogar: persona, la primera etapa es muy estable, la definición de la segunda podría cambiar cada diez años (con el nuevo censo), pero las unidades de tercera y cuarta etapa pueden cambiar en cualquier momento (nacimiento, fallecimiento, mudanza, ...)
- . Normalmente, las unidades de primera y segunda etapa se definen en términos de áreas censales u otras regiones administrativas, mientras las últimas unidades se basan en listas más recientemente actualizadas
- . En cada etapa, la selección de la muestra puede utilizar cualquier

método de muestreo, con probabilidades iguales o desiguales

... Ejemplo: Una encuesta sobre la producción agrícola necesita visitas personales para la verificación de los datos fiscales y para la observación directa de los campos. Los costos de los viajes son altos y no se puede obtener un listado completo de las fincas

- . Dentro de la provincia, hacer una lista de las 34 regiones agrícolas
- . Seleccionar 7 regiones por MASSR
- . Calcular la población total de cada una de estas 7 regiones, según datos del último censo
- . Utilizando la población como medida del tamaño, seleccionar 3 municipios dentro de cada una de las 7 regiones, con probabilidad proporcional al tamaño de cada municipio dentro de cada región
- . Para cada una de las 21 municipalidades seleccionadas, hacer una lista completa de los granjeros
- . En cada municipio, seleccionar 10 granjeros por MASSR; en el caso de un municipio con menos de 10 granjeros, tomarlos todos

¿Porqué utilizar muestreo multi-etápico?

- . Si un marco muestral apropiado para otro método no existe no se puede construir dentro del tiempo y del presupuesto disponibles
- . Puede ser factible, para cada etapa, recolectar la información necesaria para la preparación de la próxima etapa, por ejemplo datos para un segundo nivel de estratificación
- . A menudo es posible reducir los costos de viajes y de la recolección de los datos de manera importante
- . Si el efecto del diseño (coeficiente de correlación intraclase) que reduce la eficiencia de un diseño por conglomerados es muy fuerte, se puede disminuirlo con un diseño multi-etápico
- . Es posible utilizar distintos métodos de muestreo en las varias etapas, o tal vez en las distintas unidades de una misma etapa

Ventajas

- . No es necesario listar toda la población
- . Se puede eliminar una parte de las pérdidas de eficiencia asociadas con el puro muestreo por conglomerados



### Desventajas

- . Particularmente sin uno o más de las etapas utiliza muestreo con probabilidades desiguales, la estimación y más aún la estimación de la varianza pueden ser demasiado complejas

### Información necesaria para el marco muestral

- . La posibilidad de construir el marco para la próxima etapa dentro de las unidades seleccionadas en una etapa determinada, incluyendo la información necesaria para la aplicación del método de muestreo elegido

### Muestreo con probabilidades desiguales

- . El muestreo probabilístico requiere la posibilidad de calcular la probabilidad de selección de cada unidad de selección
- . No necesita que esta probabilidad sea la misma para cada unidad
- . Si se sabe que hay diferencias importantes entre las unidades con respecto a una variable que se puede llamar una medida de la importancia o del tamaño de la unidad, esta información se puede utilizar en el diseño muestral
- . Si esta medida del tamaño está más o menos directamente relacionada con las variables de interés, se puede utilizar uno de los métodos que se llaman "con probabilidad proporcional al tamaño" (PPT)
- . La probabilidad de selección de una unidad es proporcional a su tamaño, representada como una fracción del tamaño total de la población multiplicado por el número unidades que hay que seleccionar
- . La medida del tamaño puede ser el número de elementos más pequeños contenidos en la unidad, o su área geográfica, ingreso, etc.

Se han desarrollado varios métodos de muestreo con PPT. Dos de los más importantes son:

#### PPT aleatorio

- . El tamaño total de la población se calcula y a cada unidad se le asigna una parte del recorrido total que corresponde a su parte del tamaño total
- . Para cada unidad requerida en la muestra, se selecciona un número aleatorio entre 1 y el valor del tamaño total
- . La unidad cuya parte del recorrido total incluye este número está seleccionada
- . La selección se puede hacer con o sin reposición
- . Si se hace sin reposición, la parte del recorrido total que corresponde a cada unidad ya seleccionada se debe omitir para la próxima selección

ción, lo que necesita recalcular las probabilidades cada vez. Esto puede ser complicado

#### PPT sistemático

- . Se calcula y se acumulan las medidas de tamaño como en el caso anterior
- . El método básico de muestreo sistemático se aplica utilizando el tamaño total en lugar del número de unidades
- . Se divide el tamaño total por el tamaño  $n$  de la muestra para obtener un intervalo de muestreo basado en las medidas de tamaño

#### Ventajas

- . Si la correlación entre la medida del tamaño y la variable de interés es fuerte (y positiva), se mejora la precisión de las estimaciones de manera muy importante

#### Desventajas

- . Es necesario conocer el tamaño de cada unidad
- . El cálculo de las probabilidades de selección y de las estimaciones es muy complejo
- . El cálculo de las estimaciones de la varianza es mucho pero aún
- . Si hay unidades muy grandes con respecto al tamaño total, se pueden presentar problemas de selección múltiple, o puede ser necesario tratar independientemente estas unidades

#### Información necesaria para el marco muestral

- . Una lista de todas las unidades de la población con su tamaño  
o
- . Una lista de áreas geográficas con su tamaño

## RESUMEN

Hay dos tipos básicos de proceso en el diseño muestral:

Estructura de la población

- . Estratificación
- . Conglomerados

Selección de unidades

- . Muestreo probabilístico
  - . con o sin reposición
  - . probabilidades iguales
    - . MAS
    - . sistemático
  - . probabilidades desiguales
    - . PPT aleatorio
    - . PPT sistemático
- . Muestreo no probabilístico

MUESTREO PROBABILISTICO

Muestreo aleatorio simple

- . A veces difícil de seleccionar las unidades
- . Puede ser caro
- . Generalmente menor varianza (mayor precisión) que diseños multi-etápicos

Muestreo sistemático

- . Selección más sencilla que MAS
- . Problema de periodicidad

Muestreo estratificado

- . Generalmente mejora la precisión
- . Se necesita datos adecuados sobre todas las unidades, para estratificar

Muestreo por conglomerados

- . Mas barato que MAS
- . Generalmente menos preciso que MAS
- . Se simplifica la construcción del marco muestral

### Muestreo multi-etápico

- . Más barato que MAS
- . Generalmente menor precisión que MAS pero mayor que puro muestreo por conglomerados
- . Construcción del marco muestral requiere listas de unidades por etapa sólo dentro de unidades seleccionadas a la etapa anterior.

## IX PRUEBA DE HIPOTESIS E INTERVALOS DE CONFIANZA

Todo lo que se desarrolle en el presente capítulo está bajo el contexto de que la información básica que se posee, está obtenida a nivel muestral (no importando el tamaño de ella, grande o pequeño), y con el análisis de ella poder sacar conclusiones válidas, confiables acerca de lo que sucede en la población de donde provienen las muestras

## A. PRUEBAS DE HIPOTESIS PARA PROPORCIONES E INTERVALOS DE CONFIANZA

Hay muchos problemas prácticos en que la respuesta que se obtiene, o el comportamiento de la variable aleatoria, sólo admite dos valores. Ellos Generalmente un "si" o un "no", en sentido cuantitativo un "1" o un "0". Este tipo de problemas tiene la proporción "p" como parámetro de la población

Ejemplo: - Un alumno reprueba o aprueba una asignatura  
 - Un alumno repite o es promovido de curso  
 - Una dueña de casa utiliza o no utiliza un cierto producto  
 - Una persona confía o desconfía

Puede que nuestra población o universo sean todos los alumnos de la ciudad que se matriculan en Enseñanza Media. El problema a estudiar puede ser "La proporción (o porcentaje) de reprobados"

En el muestreo se puede considerar un "Éxito" una persona que reprueba un determinado curso y "fracaso" cuando ocurre la situación contraria

Un éxito (p) se registra con un UNO y el "fracaso" (q) se registra con un "cero" (0).

Por lo tanto si "Y" es una variable aleatoria cuyos posibles valores son cero o uno, entonces en una muestra de tamaño n, se obtiene una proporción muestral p.

$$\hat{p} = \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \quad \begin{matrix} 0 \leq \hat{p} \leq 1 \\ \hat{p} + \hat{q} = 1 \end{matrix}$$

Un intervalo de confianza  $(1 - \alpha)$  para la verdadera proporción p en la población, basado en la información de una muestra grande es

$$\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Ejemplo : Se toma una muestra aleatoria de 100 estudiantes de 8<sup>o</sup> años y se les consulta acerca de su preferencia por determinada asignatura (castellano). 30 contestaron afirmativamente y

$$\hat{p} = \frac{30}{100} = \frac{3}{10} = 0,3$$

$$\hat{q} = 0,7$$

Un intervalo del 90% para los alumnos de 8<sup>vo</sup> que prefieren la asignatura de Castellano es

$$(0,3) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{0,3 \times 0,7}{100}} \leq p \leq (0,3) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{0,3 \times 0,7}{100}}$$

Se toma una muestra aleatoria de 200 alumnos en el 1<sup>er</sup> año básico de una ciudad para estudiar la proporción de dislexias. En la muestra observamos 25 niños con problemas de dislexia. ¿Cuál es el intervalo de confianza del 95% para la verdadera proporción de disléxicos en 1<sup>er</sup> año básico de esa ciudad.

$$\hat{p} = \frac{25}{200} = \frac{1}{8}, \quad \hat{q} = \frac{7}{8}$$

$$\frac{1}{8} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{(0,125)(0,875)}{200}}$$

Tipos de Hipótesis

Hipótesis nula  $H_0 : P = P_0$

Hipótesis Alternativas { a)  $H_1 : P = P_0$

b)  $H_1 : P > P_0$

c)  $H_1 : P < P_0$

$$\frac{\hat{p} - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} = z_{\text{observado}}$$

## Diferencias entre Proporciones

Quando se hace una misma consulta, en problemas distintos, tenemos una proporción para cada población. Por ejemplo: se estudia la proporción de disléxicos en niños y niñas de 1<sup>er</sup> año

Se obtiene una muestra aleatoria de 150 niños (varones) y se observa que hay 15 con problemas de dislexia

Se obtiene una muestra aleatoria de 200 niñas y se observa que hay 25

$$\hat{p}_1 = \frac{15}{150} = \frac{1}{10} \quad ; \quad \hat{q}_1 = \frac{9}{10} \quad ; \quad n_1 = 150$$

$$\hat{p}_2 = \frac{25}{200} = \frac{1}{8} \quad ; \quad \hat{q}_2 = \frac{7}{8} \quad ; \quad n_2 = 200$$

Intervalos de confianza para la diferencia de proporciones

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

con  $n_1$  y  $n_2 > 30$

Dóxicas de Hipótesis

$$H_0 : P_1 - P_2 = \mathcal{J}$$

$$H_1 : P_1 - P_2 \neq \mathcal{J}$$

$$\frac{(\hat{p}_1 - \hat{p}_2) - \mathcal{J}}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = z_{\text{obs}}$$

Quando  $z_{\text{obs}}$  está fuera del intervalo

$$\left[ - z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}} \right]$$

entonces rechazamos  $H_0$

Puede también plantearse la hipótesis de que

$$H_0 : P_1 - P_2 = \mathcal{J}$$

$$H_1 : P_1 - P_2 > \mathcal{J}$$

entonces el estadístico de prueba sigue siendo

$$\frac{(\hat{p}_1 - \hat{p}_2) - \mathcal{J}}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = Z_{\text{obs}}$$

y se rechaza  $H_0$ , en favor de  $H_1$ , cuando  $Z_{\text{obs}} > Z_{1-\alpha}$

Similarmente para la hipótesis

$$H_0 : P_1 - P_2 = \mathcal{J}$$

$$H_1 : P_1 - P_2 < \mathcal{J}$$

Rechazamos  $H_0$  cuando

$$Z_{\text{obs}} < -Z_{1-\alpha}$$

### Tamaño de la muestra

Quando las variables que investigamos es del tipo de dicotómica, es decir

$$Y_1 = \begin{cases} 1 & \text{Si el sujeto posee la característica en estudio} \\ 0 & \text{Si no posee la característica} \end{cases}$$

$$\text{y } \hat{p} = \frac{\sum_{i=1}^n Y_1}{n} = \frac{x}{n} \text{ es el estimador de } P = \frac{\sum_{i=1}^N Y_1}{N}$$

y nosotros estamos interesados en saber ¿Cuál es el tamaño de la muestra que debo tomar para tener un nivel de confiabilidad deseado, aceptando un error de muestreo fijo?

$$\text{Esto nos lleva a: } n_0 = \left( \frac{Z_{1-\frac{\alpha}{2}}}{d} \right)^2 \cdot P \cdot Q$$

donde  $1 - \alpha$  es el nivel de confiabilidad deseado

$d$  : error de muestreo aceptado

$p$  : valor teórico de la proporción en la población

Casi siempre el valor de  $P$ , y por lo tanto  $Q = 1 - P$ , es una expre-



sión numérica entre cero y uno que maneje el investigador de acuerdo con:

- a) Antecedentes que él tiene del comportamiento de la proporción P en la población
- b) Resultado de una muestra piloto

El valor P · Q, es máximo cuando P = 0,5 y por lo tanto maximiza "n", para un  $t_{1-\frac{\alpha}{2}}$  y d fijos, cuando el cociente  $\frac{n_0}{N} > 0,05$  se debe calcular el tamaño muestral, y este será

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

#### Aplicaciones de la t de student

Una variable en estudio de interés es una población cualquiera, puede ser:

- Nivel de fecundidad
- Cociente intelectual
- Estatura
- Peso
- Edad
- Temperatura, etc.

Estas variables, originalmente obedecen al apellido de continuas, puesto que para cada medición que hagamos el "valor exacto" es muy difícil de obtener puesto que dependen del instrumento y la unidad de medida con que se obtienen

- a) Resulta importantísimo saber que cuando se obtiene una muestra de la población en estudio y en ella obtenemos un valor  $\bar{X}$  de la característica o variable en estudio ¿cuál es el rango en que se encuentra el verdadero valor promedio de la población, " $\mu$ ", para esa variable?

$$\bar{X} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}$$

(donde  $\mu$  es promedio en la población desconocido)

- $\bar{X}$  : promedio muestral
- s : desviación estandar muestral
- n : tamaño de la muestra

$t_{1-\frac{\alpha}{2}}$  : valor de tabla para nivel de significación  $\alpha$  y  $\nu = n - 1$   
2 grados de libertad

- b) Para el caso de dos muestras independientes es decir cuando los elementos de una de ellas no están asociados de ninguna manera con los de la otra, podemos hacer inferencias respecto de la diferencia entre dos promedios. Un problema interesante es decidir si la diferencia obtenida a nivel muestral es o no significativa. Para resolver este problema recurrimos al estadístico

$$t_{\text{obs}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{n_1 S_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right) \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}}$$

que nos sirve para decidir si es significativa la diferencia observada a nivel muestral, con nivel de confiabilidad  $(1 - \alpha)$ , de cualquiera de las siguientes tres hipótesis planteadas

1.  $H_0 : \mu_1 - \mu_2 = \delta$   
 $H_1 : \mu_1 - \mu_2 \neq \delta$  a nivel de significación  $\alpha$

$$\text{bajo } H_0 : t_{\text{obs}} = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\left(\frac{n_1 S_1^2 + n_2 s_1^2}{n_1 + n_2 - 2}\right) \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}}$$

y decimos que la diferencia observada a nivel muestral es significativa si  $t_{\text{obs}}$  está fuera del intervalo  $(-t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}})$

con  $\nu = n_1 + n_2 - 2$  grados de libertad

2.  $H_0 : \mu_1 - \mu_2 = \delta_0$   
 $H_1 : \mu_1 - \mu_2 > \delta_0$

3.  $H_0 : \mu_1 - \mu_2 = \delta_0$   
 $H_1 : \mu_1 - \mu_2 > \delta_0$

Ejemplo: Se aplica un test que mide cociente intelectual a una muestra alterna de 26 niños de un colegio, todos de la misma edad, y se tuvo que:

$$\bar{X} = 90 \quad \text{con } s = 15$$

Deseamos saber con un 95% de confiabilidad, entre que valores se encuentra el cociente intelectual de todos los niños de esa edad en el colegio

b) El estadístico 
$$\frac{(\bar{X} - \mu) \sqrt{n-1}}{s} = t_{\text{obs}}$$

nos sirve para decidir cualquiera de las tres siguientes pruebas de hipótesis.

1.  $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$  a nivel de significación  $\alpha$

bajo  $H_0$  tenemos que

$$\frac{(\bar{X} - \mu_0) \sqrt{n-1}}{s} = t_{\text{obs}}$$

y rechazamos  $H_0$  si  $t_{\text{obs}}$  está fuera del intervalo

$$\left[ -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} \right] \quad \text{con grados de libertad } \nu = n - 1$$

2.  $H_0 : \mu = \mu_0$

a nivel  $\alpha$

$H_1 : \mu > \mu_0$

entonces  $t_{\text{obs}} = \frac{(\bar{X} - \mu_0) \sqrt{n-1}}{s}$  y se rechaza  $H_0$  cuando

$$t_{\text{obs}} > t_{1-\alpha} \quad \text{con } \nu = n - 1 \text{ grados de libertad}$$

3.  $H_0 : \mu = \mu_0$

a nivel  $\alpha$  y se rechaza  $H_0$  cuando

$H_1 : \mu < \mu_0$

$$t_{\text{obs}} = \frac{(\bar{X} - \mu_0) \sqrt{n-1}}{s}$$

es menor que  $-t_{1-\alpha}$  con  $\nu = n - 1$  grados de libertad

Un intervalo de confianza de nivel  $(1 - \alpha)$  para la verdadera diferencia entre los promedios a nivel de población está dado por

$$\begin{aligned}
 (\bar{X}_1 - \bar{X}_2) - t_{1-\frac{\alpha}{2}} \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + \\
 + t_{1-\frac{\alpha}{2}} \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}
 \end{aligned}$$

## X APLICACIONES DE LA DISTRIBUCION JI-CUADRADO

Muchas veces los resultados obtenidos de muestras no siempre coinciden o concuerdan exactamente con los resultados teóricos esperados según las reglas de probabilidad. Por ejemplo : consideraciones teóricas nos dicen que cuando una moneda correctamente fabricada se lanza 100 veces y se observa la superficie superior, deberíamos obtener " 50 caras " y " 50 sellos ", es muy raro que se obtengan exactamente estos resultados.

Supóngase que en una determinada muestra se observan una serie de posibles sucesos  $S_1, S_2, \dots, S_k$  que ocurren con frecuencia  $O_1, O_2, \dots, O_k$  respectivamente, llamadas FRECUENCIAS OBSERVADAS y que según las reglas de probabilidad se espera que ocurran con frecuencias  $e_1, e_2, \dots, e_k$  respectivamente, llamadas FRECUENCIAS TEORICAS O ESPERADAS.

### a) PRUEBA DE BONDAD DE AJUSTE

La prueba ji-cuadrado (chi-cuadrado) puede ser empleada para determinar de que forma, distribuciones muestrales o tablas de frecuencias pueden ser ajustadas o llevadas al comportamiento de distribuciones teóricas de probabilidad, como el caso de una distribución normal, binomial, Poisson, etc., y con algún grado de confiabilidad decidir si se ajustan los datos empíricos o muestrales a alguna (s) de esa (s) distribución (es).

Para tal efecto se plantean dos hipótesis de trabajo :  $H_0$  y  $H_1$ , llamadas hipótesis nula o de prueba e hipótesis alternativa.

Para resolver con cual de ellas tendremos que quedarnos, en base del análisis de los resultados obtenidos, seguiremos al siguiente procedimiento.

#### a.1. Planteamiento de las hipótesis

- i) Hipótesis Nula  $H_0$  : Los datos muestrales ajustados a una determinada distribución de probabilidades, efectivamente proviene de ella.
- ii) Hipótesis alterna  $H_1$  : Los datos muestrales no provienen de la distribución de probabilidades teórica a que se ajustaron

a.2. Nivel de Significación :  $\alpha$  ( Probabilidad de error ) Nivel de confianza  $1 - \alpha$ . Si el nivel de confianza es de 0,90, significa "probable " 0,95 significa " muy probable " 0,99 " altamente probable "

a.3. Estadístico a usar

$$\chi^2_{\text{observ.}} = \sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i}$$

donde  $O_i$  : frecuencia derivada en la clase  $i$ ésima

$e_i$  : frecuencia esperada en la clase  $i$ ésima

$k$  : número de clases, sucesos, intervalos, etc.

### Ejemplo

El número de alumnos por semana que sufren algún tipo de accidente, en un gran sector escolar, durante las 36 semanas del período escolar es la siguiente :

X :	Nº de alumnos accidentados	0	1	2	3	4	total
	Nº de semanas frecuencia	6	8	10	6	6	

Los datos anteriores nos indican que el número de accidentes escolares por semana tiene una distribución de Poisson, con una significación del 5% ?

### Solución

Las hipótesis a plantearse son :

$H_0$  : El número de accidentes escolares por semana sigue una distribución de Poisson

$H_1$  : El número de accidentes escolares semanales no proviene de una distribución de Poisson

La distribución de Poisson tiene la siguiente expresión Matemática

$$P(X = k) = \frac{X^k e^{-\lambda}}{k!}$$

donde  $X$  es la variable aleatoria

$k$  = el valor que toma la variable aleatoria

$\lambda$  = promedio

$e$  = constante = 2,71828 ....

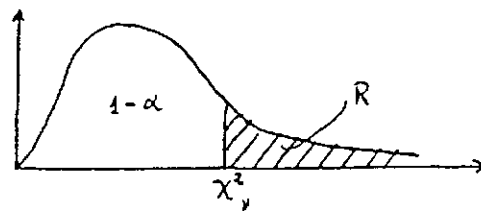
como no conocemos  $\lambda$ , buscaremos su valor haciendo una estimación basado en los resultados observados ( muestrales )

$$\hat{\lambda} = \frac{0(6) + 1(8) + 2(10) + 3(6) + 4(6) + 5(6)}{6 + 8 + 10 + 6 + 6} = \frac{82}{36} = 2,277 \approx 2,28$$

Procedemos a calcular con este valor  $\hat{\lambda}$ , las probabilidades esperadas, que multiplicadas por el total de semanas, nos entregan las "frecuencias esperadas"

a.4. Región de rechazo

$$R = \{ \chi^2 / \chi^2_{\text{observ.}} > \chi^2_{\text{crítico}} \}$$



Buscamos en la tabla de la distribución  $\chi^2$ , el valor que llamaremos  $\chi^2_{\text{crítico}}$  para lo cual debemos conocer "LOS GRADOS DE LIBERTAD  $\nu$ " con que estamos trabajando y el nivel de significación  $\alpha$  de nuestra prueba de la hipótesis  $H_0$

$$\nu = k - 1 - m$$

$k$  = número de clases, sucesos, intervalos, etc.

$m$  = número de parámetros de la distribución

a.5. Regla de decisión

Si  $\chi^2_{\text{obs.}} > \chi^2_{\text{crítico}}$  entonces rechazamos  $H_0$  para quedarnos con la hipótesis  $H_1$ , es decir que las frecuencias observadas difieren significativamente de las frecuencias esperadas. O lo que es lo mismo, los datos muestrales no provienen o no se ajustan a la distribución teórica ( con nivel de significación  $\alpha$  )

Si  $\chi^2_{\text{obs.}} \leq \chi^2_{\text{crítico}}$  entonces no rechazamos  $H_0$ , es decir, los datos muestrales ajustan o provienen de esa distribución ( con de significación  $\alpha$  )

$$\hat{P}_1 = P(X=0) = \frac{(2,28)^0 \cdot e^{-2,28}}{0!} = e^{-2,28}$$

PROBABILIDADES ESPERADAS	FRECUENCIAS ESPERADAS
$\hat{P}_1 = P(X=0) = \frac{(2.28)^0 e^{-2.28}}{0!} = e^{-2.28} = 0.102284$	$\hat{f}_1 = \hat{P}_1 \times 36 = 3.68$
$\hat{P}_2 = P(X=1) = \frac{(2.28)^1 e^{-2.28}}{1!} = 0.233208$	$\hat{f}_2 = \hat{P}_2 \times 36 = 8.4$
$\hat{P}_3 = P(X=2) = \frac{(2.28)^2 e^{-2.28}}{2!} = 0.2658571$	$\hat{f}_3 = \hat{P}_3 \times 36 = 9.57$
$\hat{P}_4 = P(X=3) = \frac{(2.28)^3 e^{-2.28}}{3!} = 0.2020514$	$\hat{f}_4 = \hat{P}_4 \times 36 = 7.27$
$\hat{P}_5 = P(X=4) = \frac{(2.28)^4 e^{-2.28}}{4!} = 0.1151693$	$\hat{f}_5 = \hat{P}_5 \times 36 = 4.15$

Luego procedemos a confeccionar la siguiente tabla

X	FRECUENCIAS OBSERVADA	FRECUENCIAS ESPERADA	$(O_i - e_i)^2 / e_i$
0	6	3.68	1.46261
1	8	8.4	0.01905
2	10	9.57	0.01932
3	6	7.27	0.22186
4	6	4.15	0.624

$$\sum_{i=1}^5 \frac{(O_i - e_i)^2}{e_i} = 2.34715$$

Luego debemos buscar el punto crítico o el  $\chi^2$  crítico :

Determinamos  $V = k - 1 - m$  grados de libertad

$k = 5$  puesto que hay cinco intervalos

$m = 1$  puesto que la distribución tiene un parámetro estimado  $\lambda$

Entonces  $V = k - 1 - m = 5 - 1 - 1 = 3$



El nivel de significación  $\alpha = 0.05$

$$\chi^2_{\nu=3} (\alpha = 0.05) = 7.81$$

Por lo tanto, para decidir la hipótesis con que nos quedaremos tenemos que comparar el  $\chi^2_{\text{obsv.}}$  con el  $\chi^2_{\text{crítico}}$

$$\chi^2_{\text{obsv.}} = 2.55$$

$$\chi^2_{\text{obsv.}} = 7.81$$

$$\chi^2_{\text{obsv.}} < \chi^2_{\text{crítico}}, \text{ entonces}$$

No rechazamos  $H_0$ ; es decir los datos provienen o ajustan a una distribución de Poisson, con una significación  $\alpha = 0.05$

Desarrollamos otro ejemplo de aplicación de la  $\chi^2$  a la prueba de bondad del ajuste.

La puntuación final en Matemáticas de 80 estudiantes de los cursos de cuartos medios de un Liceo, en escala de 0 a 100 puntos se registró en la siguiente tabla

INTERVALOS	$f_i$
40.5 - 50.5	12
50.5 - 60.5	20
60.5 - 70.5	17
70.5 - 80.5	18
80.5 - 80.5	8
90.5 - 100.5	5

El investigador desea saber si los rendimientos de los cursos es probable que se ajusten a una distribución normal de probabilidades.

Esta distribución nos exige conocer los parámetros;

$\mu$  = promedio de la población

$\sigma^2$  = varianza de la población

Ellos son estimados mediante el promedio aritmético  $\bar{x}$  de la muestra y la varianza muestral  $s^2$  respectivamente.

En este caso

$$\bar{X} = 66.125 \quad S = 14.26$$

La distribución normal de probabilidades tiene como función matemática :

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}}$$

Nosotros no ocuparemos esta expresión algebraica, si no que con el promedio  $\bar{X}$  y la desviación estándar  $S$  obtenidas de los datos recopilados, procedemos a la " estandarización de los límites reales de los intervalos " lo que nos permitirá calcular las probabilidades esperadas mediante el manejo de una tabla estadística llamada : Distribución normal estándar o típica.

Ejemplo de estandarización :

$$z_i = \frac{\text{límite real del intervalo} - \bar{X}}{S}$$

$$z_1 = \frac{40.5 - 66.125}{14.26} = -1.8$$

Construimos la siguiente tabla

INTERVALOS REALES	INTERVALOS ESTANDARIZADOS	PROBABILIDADES ESPERADAS	FRECUENCIAS ESPERADAS
40.5 - 50.5	-1.80 -1.1	0.0976	0.0976x30 = 2.9
50.5 - 60.5	-1.10 -0.39	0.2148	0.2148x30 = 6.4
60.5 - 70.5	-0.39 0.31	0.2734	0.2734x30 = 8.2
70.5 - 80.5	0.31 1.01	0.2221	0.2221x30 = 6.7
80.5 - 90.5	1.01 1.71	0.1126	0.1126x30 = 3.4
90.5 - 100.5	1.71 2.41	0.0356	0.0356x30 = 1.1

Seguidamente colocamos dos columnas, que llevan las frecuencias observadas, frecuencias esperadas y nos permiten calcular

$$\chi^2_{\text{observ.}} = \sum_{i=1}^6 \frac{(O_i - e_i)^2}{e_i}$$

FRECUENCIAS OBSERVADAS	FRECUENCIAS ESPERADAS	$(O_i - e_i)^2 / e_i$
12	8	2
20	17	0.5294
17	22	1.1364
18	18	0
8	9	0.1111
5	3	1.3333

Y el  $\chi^2_{\text{obsrv.}} = 5.1102$

Por lo tanto procedemos a determinar el  $\chi^2_{\text{crítico}}$

con  $v = k - 1 - m = 3$

con  $k = 6$  por haber 6 intervalos

$m = 2$  por haber estimado dos parámetros

con  $\alpha = 0.1$

Entonces  $\chi^2_{\text{crítico}} = 6.25$

Como  $\chi^2_{\text{obsrv.}} = 5.11 < \chi^2_{\text{crítico}} = 6.25$

No rechazamos  $H_0$  y podemos decir con una significación  $\alpha = 0.1$

que los rendimientos en matemática de los dos cuartos años del Liceo siguen una distribución normal de probabilidades.

Otra distribución de probabilidades importante en el ámbito social educativo es la distribución de probabilidades Binomial cuya expresión algebraica es :

$$B(X = k ; p ; n) = C_k^n p^k (1-p)^{n-k}$$

donde

$X$  variable aleatoria que puede tomar cualquier valor natural entre 0 y  $n$  ( $0 \leq k \leq n$ )

$p$ : probabilidad de éxito del suceso

$C_k^n$  : combinaciones de  $k$  objetos sobre los  $n$  disponibles

$$C_k^n = \frac{n!}{(n-k)! k!}$$

Ejemplo  $C_3^5 = \frac{5!}{(5-3)! 3!} = \frac{(5 \cdot 4 \cdot 3) \times (2 \cdot 1)}{(2 \cdot 1) (3 \cdot 2 \cdot 1)} = \frac{120}{2 \times 6} = 10$

Ejemplo

En una población hay 100 familias que tienen cinco hijos cada una. Una investigación es realizada a objeto de saber cuantos hijos varones hay en ellas o los resultados son

N°DE HIJOS VARONES (k)	FRECUENCIAS OBSERVADAS ( N° DE FAMILIAS )
0	38
1	144
2	342
3	287
4	164
5	25

Nos preguntamos si estos recolectados provienen de una distribución binomial.

Para poder ajustar a estos datos a una distribución binomial deberos calcular el promedio aritmético

$$\bar{X} = 2.47 \quad \text{puesto que para este caso}$$

$$\bar{X} = n p \quad \text{y nos permite estimar "p"}$$

$$\hat{p} = \frac{\bar{X}}{n} = \frac{2.47}{5} = 0.494 \quad \text{que para objeto de nuestros cálculos}$$

los lo aproximaremos a  $\hat{p} = 0.5$  y con el determinar : las probabilidades esperadas y las frecuencias esperadas.

Por ejemplo

$$P (X=2; p=0.5; n=5) = C_2^5 (0.5)^2 (1-0.5)^{5-2}$$

$$C_2^5 = \frac{5!}{(5-2)! 2!} = \frac{120}{6 \cdot 2} = 10$$

$$\text{Entonces } P (X=2; p=0.5; n=5) = 10 \times (0.25) (0.125) = 0.3125$$

N° DE HIJOS V. (k)	PROBABILIDADES	ESPERADAS	FRECUENCIAS ESPERADAS
0	$C_0^5 (0.5)^0 (0.5)^{5-0} = (1) \times (1) \times (0.03125) = 0.03125$		$0.03125 \times 1000 = 31$
1	$C_1^5 (0.5)^1 (0.5)^{5-1} = 5 \times (0.5) \times (0.0625) = 0.15625$		$0.15625 \times 1000 = 156$
2	$C_2^5 (0.5)^2 (0.5)^{5-2} = 10 \times (0.25) \times (0.125) = 0.3125$		$0.3125 \times 1000 = 312$
3	$C_3^5 (0.5)^3 (0.5)^{5-3} = 10 \times (0.125) \times (0.25) = 0.3125$		$0.3125 \times 1000 = 312$
4	$C_4^5 (0.5)^4 (0.5)^{5-4} = 5 \times (0.0625) \times (0.5) = 0.15625$		$0.15625 \times 1000 = 156$
5	$C_5^5 (0.5)^5 (0.5)^{5-5} = 1 \times (0.03125) \times 1 = 0.03125$		$0.03125 \times 1000 = 31$

Seguidamente procedemos a colocar las columnas

FRECUENCIAS OBSERVADA	FRECUENCIAS ESPERADAS	$(o_i - e_i)^2 / e_i$
38	31	1.5806
144	156	0.9231
342	312	2.8846
287	312	2.0032
164	156	0.4203
25	31	1.1613

y el  $\chi^2_{\text{observ.}} = 8.9631$

Seguidamente determinamos el  $\chi^2_{\text{crítico}}$

$$\nu = k - 1 - m = 4$$

puesto que  $k = 6$  intervalos

$m = 1$ , por haber estimado un parámetro

y con  $\alpha = 0.1$

$$\chi^2_{\text{crítico}} = 7.78$$

La regla de decisión la obtenemos al comparar  $\chi^2_{\text{observ.}}$  con el

$\chi^2_{\text{crítico}}$

Como  $\chi^2_{\text{observ.}} = 8.96 > \chi^2_{\text{crítico}} = 7.78$

Rechazamos  $H_0$ , en favor de  $H_1$ , es decir los datos muestrales no provienen de una distribución binomial con un nivel de significación  $\alpha = 0.1$

Sin embargo si  $\alpha = 0.05$  y  $\nu = 4$ , el  $\chi^2_{\text{crítico}} = 9.49$  y en este caso no rechazamos  $H_0$ , y debemos concluir, que con ese nivel de significación, los datos muestrales nos indicarían que el número de hijos varones en familias de cinco hijos se ajusta a una distribución binomial.

b) Intervalo de confianza  $(1-\alpha)$  para la varianza de la población

Es muy importante determinar un intervalo de confianza para la varianza de la población cuando se tiene una muestra de " n elementos " de dicha población.

La teoría de Inferencia Estadística nos señala que

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

además

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

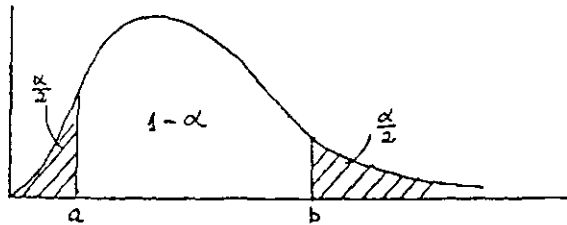
$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Por lo tanto

$$\frac{n \left( \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right)}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$n \frac{s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

En una tabla  $\chi^2$  y con  $\nu = n - 1$  grados de libertad, determinamos puntos críticos "a" y "b", para un nivel de confiabilidad  $1 - \alpha$



y en el intervalo de confianza  $1 - \alpha$  para  $\sigma^2$  es

$$\frac{ns^2}{b} \leq \sigma^2 \leq \frac{ns^2}{a}$$

o también

$$\frac{(n-1) \hat{S}^2}{b} \leq \sigma^2 \leq \frac{(n-1) \hat{S}^2}{a}$$

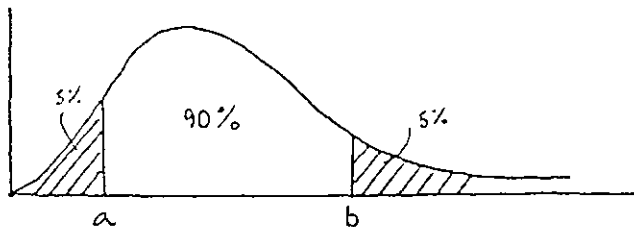
Cuando el tamaño de la muestra es grande, ambas expresiones anteriores tienden a coincidir

### Ejemplo

En una muestra de 10 observaciones se encuentra que la varianza  $S^2$  es igual a 15. Hallar los límites de confianza del 90% para  $\sigma^2$

### Desarrollo

En una tabla  $\chi^2$  ubicamos la línea  $\nu = n - 1 = 9$  grados de libertad para determinar "a" y "b"



$$a = \chi^2_{9(0.05)} = 3.33$$

$$b = \chi^2_{9(0.95)} = 16.9$$

y el intervalo pedido es

$$\frac{10 \cdot 15}{16.9} \leq \sigma^2 \leq \frac{10 \cdot 15}{3.33}$$

$$8,876 \leq \sigma^2 \leq 45,045$$

c) Contraste de hipótesis respecto de una sola varianza

Supóngase que interese determinar si una varianza de población ha cambiado con el tiempo, o si la varianza de una distribución es diferente de cierto valor fijado. Para este fin, se contrasta la hipótesis de nulidad

$$H_0 : \sigma^2 = \sigma_0^2$$

donde  $\sigma_0^2$  es la varianza conocida o algún valor hipotético.

La hipótesis  $H_1$  depende del problema que se considere.

Si  $X$  es una variable aleatoria que se distribuye Normal  $(\mu; \sigma^2)$  y si una muestra de tamaño  $n$  da  $X$  de una varianza  $S^2$ , entonces dado el nivel de significación, el estadígrafo de contraste es

$$\frac{n S^2}{\sigma_0^2}$$

Resumen : Contraste de  $H_0 : \sigma^2 = \sigma_0^2$  con

$H_1$  cuando  $X \sim N(\mu, \sigma^2)$  y

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

$H_1$	Regla de decisión : Rechazar $H_0$ si donde a nivel $\alpha$ y $\nu = n-1$
$\sigma^2 > \sigma_0^2$	$\frac{n S^2}{\sigma_0^2} \geq c_1 = \chi^2_{(\nu, 1-\alpha)}$
$\sigma^2 < \sigma_0^2$	$\frac{n S^2}{\sigma_0^2} \leq c_2 = \chi^2_{(\nu, \alpha)}$
$\sigma^2 \neq \sigma_0^2$	$\frac{n S^2}{\sigma_0^2} \geq c_3 = \chi^2_{(\nu, \alpha/2)}$ o también $\frac{n S^2}{\sigma_0^2} \leq c_4 = \chi^2_{(\nu; 1-\alpha/2)}$



### Ejemplo Aplicado

Se desea determinar si la varianza de los I.Q de los estudiantes de cierto distrito escolar es mayor que 250. Una muestra aleatoria de 30 estudiantes del distrito de una varianza de 275. Contrastar la hipótesis de que la varianza de los I.Q se mantiene en 250 al nivel del 5%

### Desarrollo

- Las hipótesis a contrastar o probar son :

$$H_0 : \sigma^2 = 250 \quad \text{v/s} \quad H_1 : \sigma^2 > 250$$

- Determinamos el valor del Estadígrafo de contraste

$$\chi^2_{\text{obsrv.}} = \frac{n S^2}{\sigma_0^2} \Rightarrow \chi^2_{\text{obsrv.}} = \frac{30 \times 275}{250} = 33$$

- Leemos en tabla  $\chi^2$  ( $\nu = 29$ ; 95%) = 42,6 =  $C_1$

- Decisión :

$$\text{como } \chi^2_{\text{obsrv.}} = 33 < C_1 = 42.6$$

entonces No rechazamos  $H_0$ , es decir, los resultados muestrales nos indican que se mantiene en 250 la varianza poblacional del I.Q de los estudiantes.

### c.- Contraste de Independencia

Esta prueba permite al investigador determinar si existe asociación entre dos variables con escala de medición Nominal u Ordinal.

Frecuentemente se encuentran problemas que exigen una " clasificación doble ", o " bidimensional " de los datos observados, generalmente en tablas llamadas de " doble entrada ", y se desea analizar si existe algún tipo de relación entre las dos variables.

Supongamos que se tienen "n" observaciones que se distribuyan conjuntamente con respecto a dos variables C y R; cada una de las cuales se clasifica en varias categorías, clases o intervalos.

Sean  $C_1, C_2, \dots, C_k$  las k clases de variable C

Sean  $r_1, r_2, \dots, r_t$  las t clases de Variables R

Las variables C y R son mutuamente excluyentes.

Cada observación representa la ocurrencia de un suceso conjunto  $(r_i, C_j)$  con  $i = 1, 2, \dots, t$   $j = 1, 2, \dots, k$

La muestra de n observaciones puede entonces estar distribuida en C x R casilla, de la siguiente manera

R	C				TOTAL
	C <sub>1</sub>	C <sub>2</sub>	C <sub>j</sub>	C <sub>k</sub>	
r <sub>1</sub>	f <sub>11</sub>	f <sub>12</sub> .....	f <sub>1j</sub> .....	f <sub>1k</sub>	f <sub>1.</sub>
r <sub>2</sub>	f <sub>21</sub>	f <sub>22</sub> .....	f <sub>2j</sub> .....	f <sub>2k</sub>	f <sub>2.</sub>
⋮	⋮	⋮	⋮	⋮	⋮
r <sub>i</sub>	f <sub>i1</sub>	f <sub>i2</sub> .....	f <sub>ij</sub> .....	f <sub>ik</sub>	f <sub>i.</sub>
⋮	⋮	⋮	⋮	⋮	⋮
r <sub>t</sub>	f <sub>t1</sub>	f <sub>t2</sub>	f <sub>tj</sub>	f <sub>tk</sub>	f <sub>t.</sub>
TOTAL	f . 1	f . 2	f . j	f . k	n

donde

$f_{ij}$  : frecuencia con que ha ocurrido el suceso conjunto  $(r_i, C_j)$

$f_j$  : frecuencia con que ha ocurrido el suceso  $r_i$

Debemos calcular a partir de la tabla anterior las frecuencias esperadas :  $\hat{f}$

$$\hat{f}_{ij} = \frac{(f_{.j}) (f_{i.})}{n} \quad \begin{array}{l} i = 1 \dots t \\ j = 1 \dots k \end{array}$$

La metodología a usar para el contraste de Independencia es la siguiente

i) Planteamiento de las hipótesis

$H_0$  : No hay asociación (correlación) entre ambas variables

$H_1$  : Existe asociación entre ambas variables

ii) Determinar nivel de significación con que se va a trabajar

iii) Determinar los grados de libertad

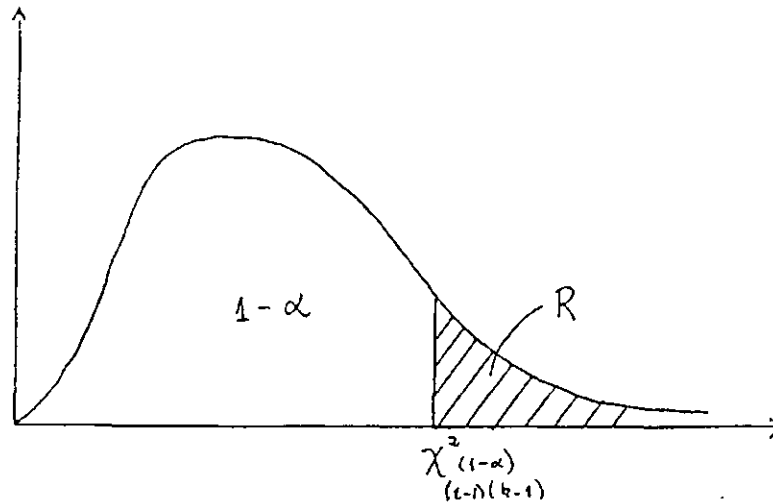
$$v = (t-1) \cdot (k-1)$$

iv) Estadístico a usar

$$\chi^2_{\text{observ.}} = \sum_{j=1}^k \sum_{i=1}^t \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

v) Región de Rechazo

$$R = \left\{ \chi^2 / \chi^2_{\text{observ.}} > \chi^2_{(t-1)(k-1)}(1-\alpha) \right\}$$



### Ejemplo

Supóngase que se desea averiguar si hay alguna asociación significativa entre el nivel de formación académica y rendimiento laboral para un grupo de 200 empleados!

El nivel de formación académica se clasifica en tres clases : Escuela Básica, Escuela Media, Escuela de Especialización.

El rendimiento en el trabajo (laboral) se clasifica como : excelente bueno, regular.

Los datos son los siguientes

RENDIMIENTO	FORMACION ACADEMICA			TOTAL
	ESCUELA BASICA	ESCUELA MEDIA	ESCUELA DE ESPECIALIZ.	
EXCELENTE	10	40	10	60
BUENO	30	30	20	80
REGULAR	10	30	20	60
TOTAL	50	100	50	200

### Desarrollo

Deberos de inmediato calcular las frecuencias esperadas  $f_{ij}$ . En es

$$\hat{f}_{11} = \hat{f}_{\text{excelente y Esc. Básica}} = \frac{f_{\cdot 1} \times f_{1\cdot}}{n} = \frac{50 \times 60}{200} = 15$$

TABLAS DE FRECUENCIAS ESPERADAS

RENDIMIENTO	FORMACION ACADEMICA		
	ESCUELA BASICA	ESCUELA MEDIA	ESCUELA DE ESPECIALIZ.
EXCELENTE	$\frac{50 \times 60}{200} = 15$	$\frac{100 \times 60}{200} = 30$	$\frac{50 \times 60}{200} = 15$
BENO	$\frac{50 \times 80}{200} = 20$	$\frac{100 \times 80}{200} = 40$	$\frac{50 \times 50}{200} = 20$
REGULAR	$\frac{50 \times 60}{200} = 15$	$\frac{100 \times 60}{200} = 30$	$\frac{50 \times 60}{200} = 15$

Como el estadístico a usar es  $\chi^2_{\text{observc.}} = \sum_{j=1}^k \sum_{i=1}^t \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$

entonces procedemos a calcular su valor

$$\begin{aligned} \chi^2_{\text{observc.}} &= \frac{(10 - 15)^2}{15} + \frac{(40 - 30)^2}{30} + \frac{(10 - 15)^2}{45} + \frac{(30 - 20)^2}{20} + \frac{(30 - 40)^2}{40} \\ &+ \frac{(20 - 20)^2}{20} + \frac{(10 - 15)^2}{15} + \frac{(30 - 30)^2}{30} + \frac{(20 - 15)^2}{15} = 17.5 \end{aligned}$$

Fijando un nivel de significación del 1%, o lo que es lo mismo, un 99% de confianza; y  $\gamma = (3-1)(3-1) = 4$  g.l.  $\chi^2_{\text{crítico}} = 13,3$

Nuestra decisión debe ser la de rechazar  $H_0$ ; puesto que

$$\chi^2_{\text{observc.}} = 17.5 > \chi^2_{\text{crítico}} = 13.3$$

Es decir, existe asociación entre ambas variables, lo que implicaría que la formación académica y el rendimiento laboral está relacionado.

#### Observación 1

Podemos también llegar a calcular un coeficiente de correlación ( $r$ ) dado que la categorización o niveles en que fueron medidas las variables son atributos y además es una tabla cuadrada de  $k \times k$  o también  $t \times t$ ,

$$r = \sqrt{\frac{\chi^2_{\text{observc.}}}{N(k-1)}} \Rightarrow r = \sqrt{\frac{17.5}{200(3-1)}} = 0,2092$$

$$\Rightarrow r = 20,92\%$$

y además, siempre para este tipo de tablas de asociación:  $-1 \leq r \leq 1$

Por lo tanto,  $r$  : se denomina correlación de atributos.

### Observación 2

Cuando la tabla bidimensional o de contingencia, etc. tiene dos filas y dos columnas ( $k = t = 2$ ) y presenta la siguiente estructura

	I	II	TOTAL
A	$a_1$	$a_2$	$N_A$
B	$b_1$	$b_2$	$N_B$
TOTALES	$N_1$	$N_2$	$N$

para calcular el  $\chi^2_{\text{observc.}}$ , debe utilizarse una fórmula especial que contempla una corrección llamada corrección de Yates

$$\chi^2_{\text{observc. corregido}} = \frac{N \left( |a_1 b_2 - a_2 b_1| - \frac{1}{2} N \right)^2}{N_1 N_2 N_A N_B}$$

### Ejemplo

El número de estudiantes de cada una de dos clases A y B que aprobó y reprobó un examen determinado es el siguiente

	APROBARON	REPROBARON
CLASE A	72	17
CLASE B	64	23

Mediante un nivel de significación  $\alpha = 0.05$  ensaye la hipótesis de que no hay diferencia entre las dos clases

Desarrollo

Las hipótesis que se plantean para ensayar son

$H_0$  : No hay diferencias significativas entre las dos clases

$H_1$  : Existen diferencias significativas

Calculamos el  $\chi^2_{\text{observado}}$  con la corrección de Yates

$$N_1 = 72 + 64 = 136$$

$$N_2 = 17 + 23 = 40$$

$$N_A = 72 + 17 = 89$$

$$N_B = 64 + 23 = 87$$

$$\begin{aligned} |a_1b_2 - a_2b_1| &= |72 \cdot 23 - 17 \cdot 64| = |1656 - 1088| \\ &= |568| = 568 \end{aligned}$$

$$\chi^2_{\text{observ. corr}} = \frac{176 \left( 568 - \frac{1}{2} \cdot 176 \right)^2}{136 \times 40 \times 89 \times 87} = 0.9627$$

Determinamos  $\chi^2_{\text{crítico}}$  a  $\alpha = 0,05$  y  $\nu = (2 - 1)(2 - 1) = 1$

$$\chi^2_{\nu=1} (1 - \alpha) = 3.84$$

y como  $\chi^2_{\text{observ.}} = 0,9627 < \chi^2_{\text{crítico}} = 3.84$

No podemos rechazar  $H_0$

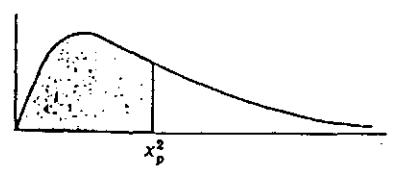
El coeficiente de correlación de atributos es

$$r = \sqrt{\frac{0.9627}{176(2-1)}} = 0.0523 \approx 5,23\%$$

APENDICES

Apéndice I

PERCENTILES ( $\chi^2_p$ )  
DE LA  
DISTRIBUCION CHI-CUADRADO  
CON  $v$  GRADOS DE LIBERTAD  
(AREA SOMBREADA =  $p$ )



$v$	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.75}$	$\chi^2_{0.50}$	$\chi^2_{0.25}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.01}$	$\chi^2_{0.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	0.455	0.102	0.0158	0.0039	0.0010	0.0002	0.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	0.575	0.211	0.103	0.0506	0.0201	0.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	0.584	0.352	0.216	0.115	0.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	0.711	0.484	0.297	0.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	0.831	0.554	0.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	0.872	0.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	0.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	166.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

Procedencia: Catherine M. Thompson, *Table of percentage points of the  $\chi^2$  distribution*, *Biometrika*, Vol. 32 (1941), con permiso de los autores y editores.

## APÉNDICES

Apéndice II.

PERCENTILES ( $t_p$ )  
DE LA  
DISTRIBUCIÓN  $t$  DE STUDENT  
CON  $v$  GRADOS DE LIBERTAD  
(ÁREA SOMBRADA =  $p$ )



$v$	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,55}$
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	0.727	0.325	0.158
2	9.92	6.96	4.30	2.92	1.89	1.061	0.816	0.617	0.289	0.142
3	5.84	4.54	3.18	2.35	1.64	0.978	0.765	0.584	0.277	0.137
4	4.60	3.75	2.78	2.13	1.53	0.941	0.741	0.569	0.271	0.134
5	4.03	3.36	2.57	2.02	1.48	0.920	0.727	0.559	0.267	0.132
6	3.71	3.14	2.45	1.94	1.44	0.906	0.718	0.553	0.265	0.131
7	3.50	3.00	2.36	1.90	1.42	0.896	0.711	0.549	0.263	0.130
8	3.36	2.90	2.31	1.86	1.40	0.889	0.706	0.546	0.262	0.130
9	3.25	2.82	2.26	1.83	1.38	0.883	0.703	0.543	0.261	0.129
10	3.17	2.76	2.23	1.81	1.37	0.879	0.700	0.542	0.260	0.129
11	3.11	2.72	2.20	1.80	1.36	0.876	0.697	0.540	0.260	0.129
12	3.06	2.68	2.18	1.78	1.36	0.873	0.695	0.539	0.259	0.128
13	3.01	2.65	2.16	1.77	1.35	0.870	0.694	0.538	0.259	0.128
14	2.98	2.62	2.14	1.76	1.34	0.868	0.692	0.537	0.258	0.128
15	2.95	2.60	2.13	1.75	1.34	0.866	0.691	0.536	0.258	0.128
16	2.92	2.58	2.12	1.75	1.34	0.865	0.690	0.535	0.258	0.128
17	2.90	2.57	2.11	1.74	1.33	0.863	0.689	0.534	0.257	0.128
18	2.88	2.55	2.10	1.73	1.33	0.862	0.688	0.534	0.257	0.127
19	2.86	2.54	2.09	1.73	1.33	0.861	0.688	0.533	0.257	0.127
20	2.84	2.53	2.09	1.72	1.32	0.860	0.687	0.533	0.257	0.127
21	2.83	2.52	2.08	1.72	1.32	0.859	0.686	0.532	0.257	0.127
22	2.82	2.51	2.07	1.72	1.32	0.858	0.686	0.532	0.256	0.127
23	2.81	2.50	2.07	1.71	1.32	0.858	0.685	0.532	0.256	0.127
24	2.80	2.49	2.06	1.71	1.32	0.857	0.685	0.531	0.256	0.127
25	2.79	2.48	2.06	1.71	1.32	0.856	0.684	0.531	0.256	0.127
26	2.78	2.48	2.06	1.71	1.32	0.856	0.684	0.531	0.256	0.127
27	2.77	2.47	2.05	1.70	1.31	0.855	0.684	0.531	0.256	0.127
28	2.76	2.47	2.05	1.70	1.31	0.855	0.683	0.530	0.256	0.127
29	2.76	2.46	2.04	1.70	1.31	0.854	0.683	0.530	0.256	0.127
30	2.75	2.46	2.04	1.70	1.31	0.854	0.683	0.530	0.256	0.127
40	2.70	2.42	2.02	1.68	1.30	0.851	0.681	0.529	0.255	0.126
60	2.66	2.39	2.00	1.67	1.30	0.848	0.679	0.527	0.254	0.126
120	2.62	2.36	1.98	1.66	1.29	0.845	0.677	0.526	0.254	0.126
$\infty$	2.58	2.33	1.96	1.645	1.28	0.842	0.674	0.524	0.253	0.126

Procedencia: R. A. Fisher y F. Yates. *Statistical Tables for Biological, Agricultural and Medical Research* (5.ª edición). Tabla III. Oliver and Boyd Ltd., Edimburgo, con permiso de los autores y editores.



### Apéndice III

Tabla I. Valores de la función distribución normal estándar\*

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = P(Z \leq z)$$

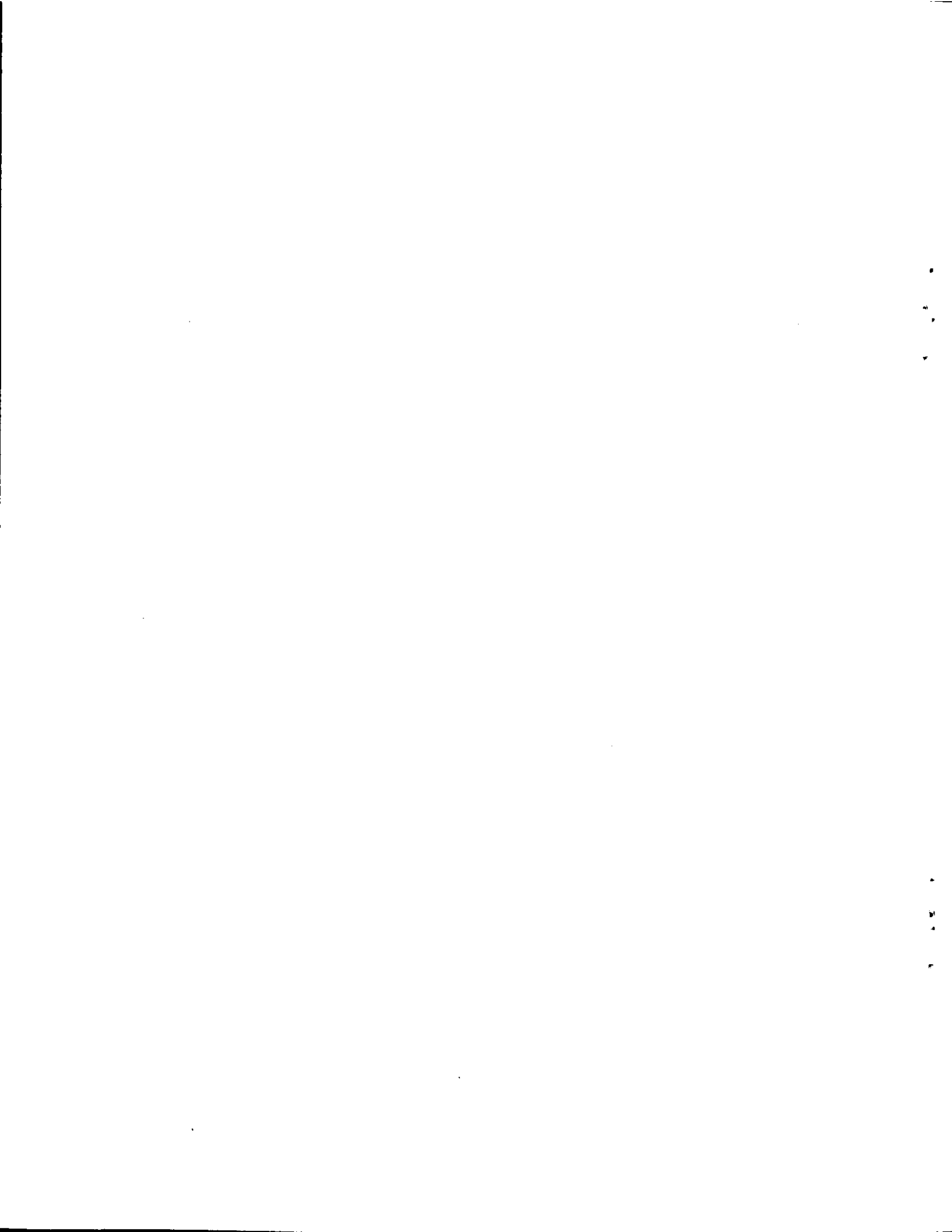
z	0	1	2	3	4	5	6	7	8	9
-3.0	0.0013	0.0010	0.0007	0.0005	0.0003	0.0002	0.0002	0.0001	0.0001	0.0000
-2.9	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0238	0.0233
-1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0300	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0570	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

\* B. W. Lindgren, *Statistical Theory*. New York: The Macmillan Co., 1960.

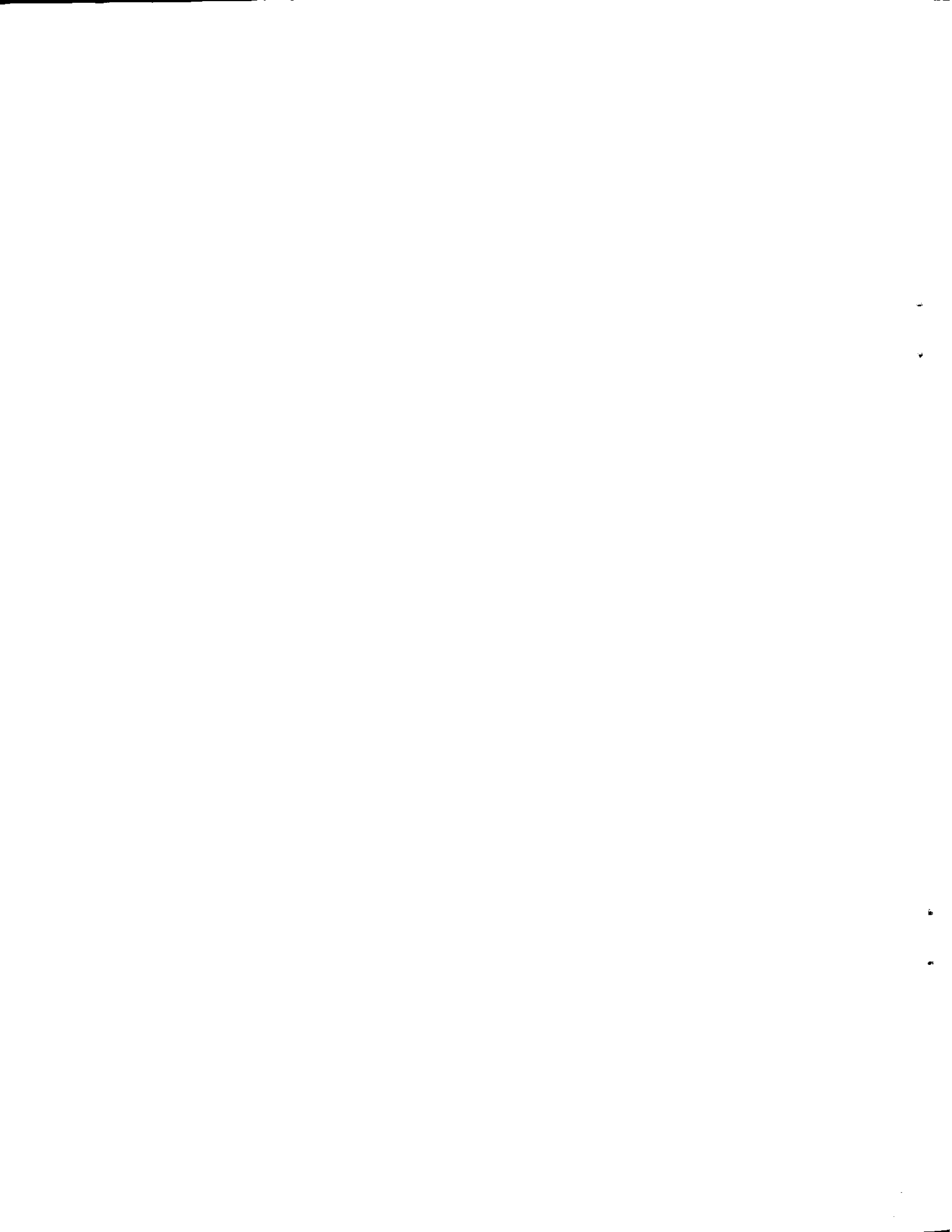
TABLA 1. (Continuación)

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = P(Z \leq z)$$

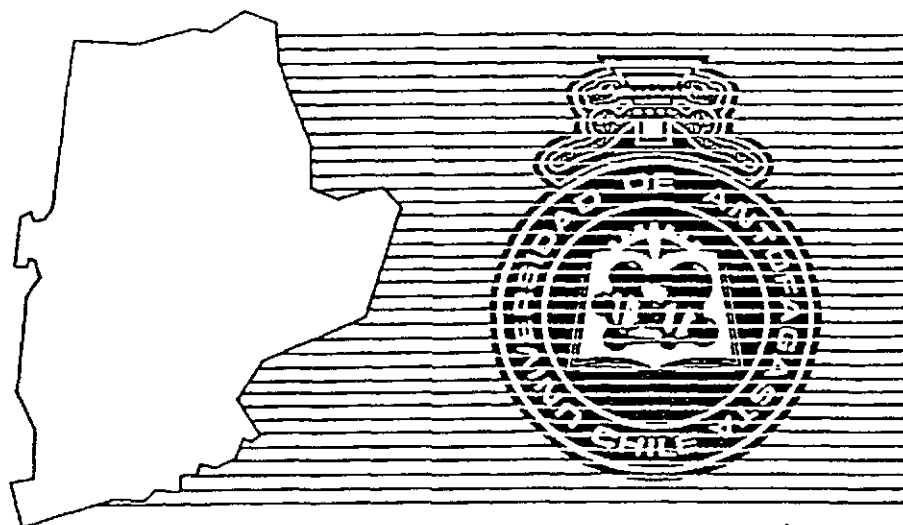
z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5595	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9795	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000







**UNIVERSIDAD DE ANTOFAGASTA**  
FACULTAD DE CIENCIAS BASICAS  
FACULTAD DE INGENIERIA



**CURSO :** TALLER DE APLICACION DE LAS TECNICAS  
DE CONTROL DE CALIDAD

**RELATOR :** SR. MAGISTER RENE MALUENDA M.

**ANTOFAGASTA - CHILE**



## PARA EMPEZAR

Cuando te preparas para ir de viaje a una ciudad que no conoces, empiezas mirando mapas y libros. Así determinas la mejor forma de llegar a la ciudad, qué posibilidades de clientes hay en ella, quiénes son las personas en cada una de las empresas que puedan pasar a formar parte de la clientela, y la mejor manera de salir de la ciudad.

Prepararse para usar STATGRAPHICS no se diferencia de prepararse para un viaje de negocios. Tienes que aprender :

- dónde están las teclas en el teclado del ordenador que uses con STATGRAPHICS.
- cómo poner en marcha STATGRAPHICS.
- cómo están organizados los diferentes medios estadísticos de STATGRAPHICS.
- cómo tener acceso a los muchos medios estadísticos que ofrece STATGRAPHICS.
- cómo se comunica contigo STATGRAPHICS.
- cómo usar variables y ficheros para almacenar los datos de STATGRAPHICS.
- cómo salir de STATGRAPHICS.

Este capítulo te explica todo esto con detalle para que puedas dar los pasos básicos necesarios para usar STATGRAPHICS.

## EL TECLADO DEL ORDENADOR.

Incluso para poder hacer las operaciones más sencillas en STATGRAPHICS, necesitas saber algo sobre algunas teclas importantes en el teclado del ordenador (por ejemplo, la tecla ENTER, las teclas de función, y las teclas de control del cursor)- lo que hacen y dónde se encuentran. Las figuras 2-1, 2-2, y 2-3 muestran tres de los teclados más comunes e indican la posición de estas teclas importantes.



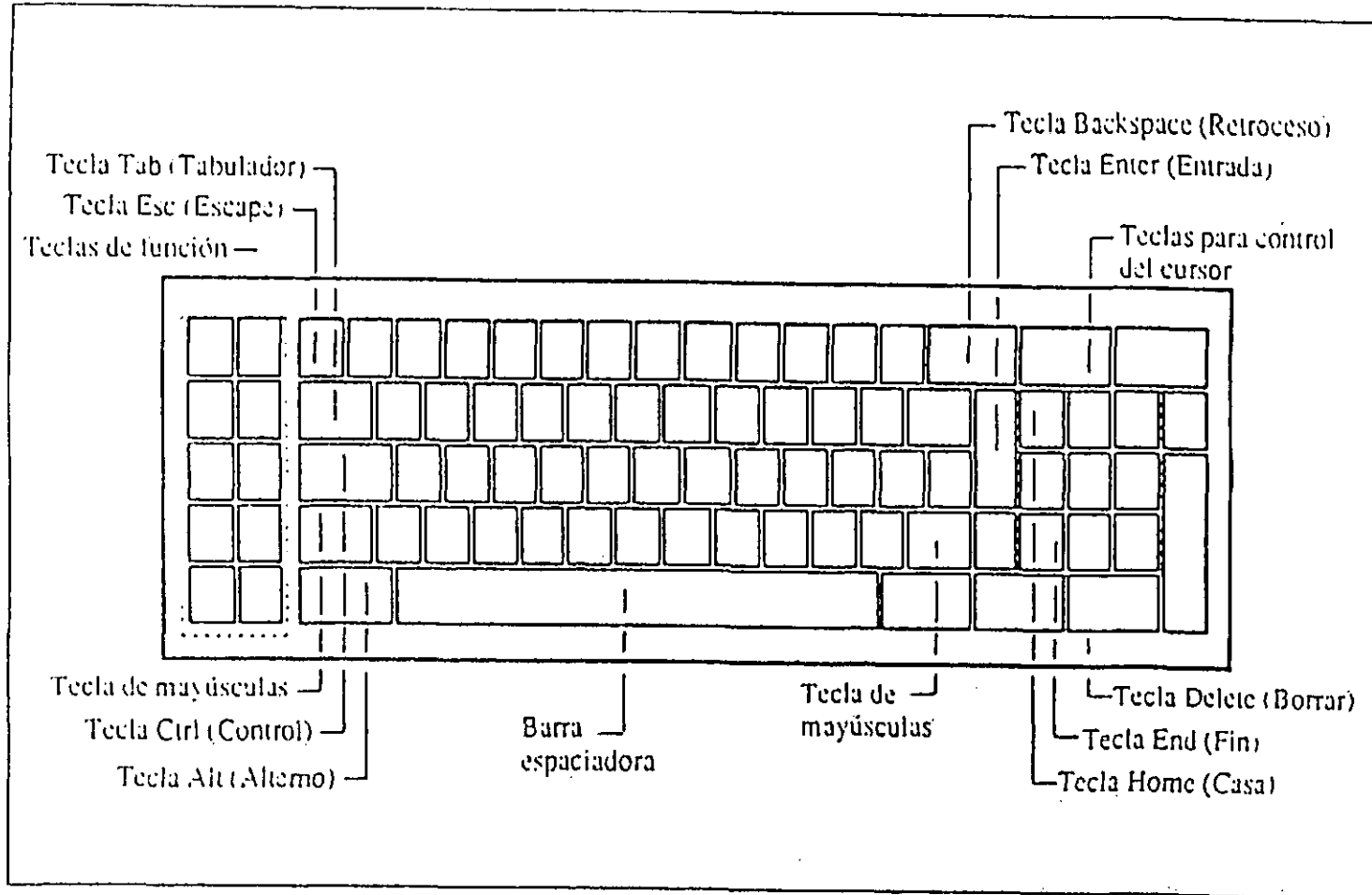


Figura 2-1. Teclado del ordenador personal de IBM

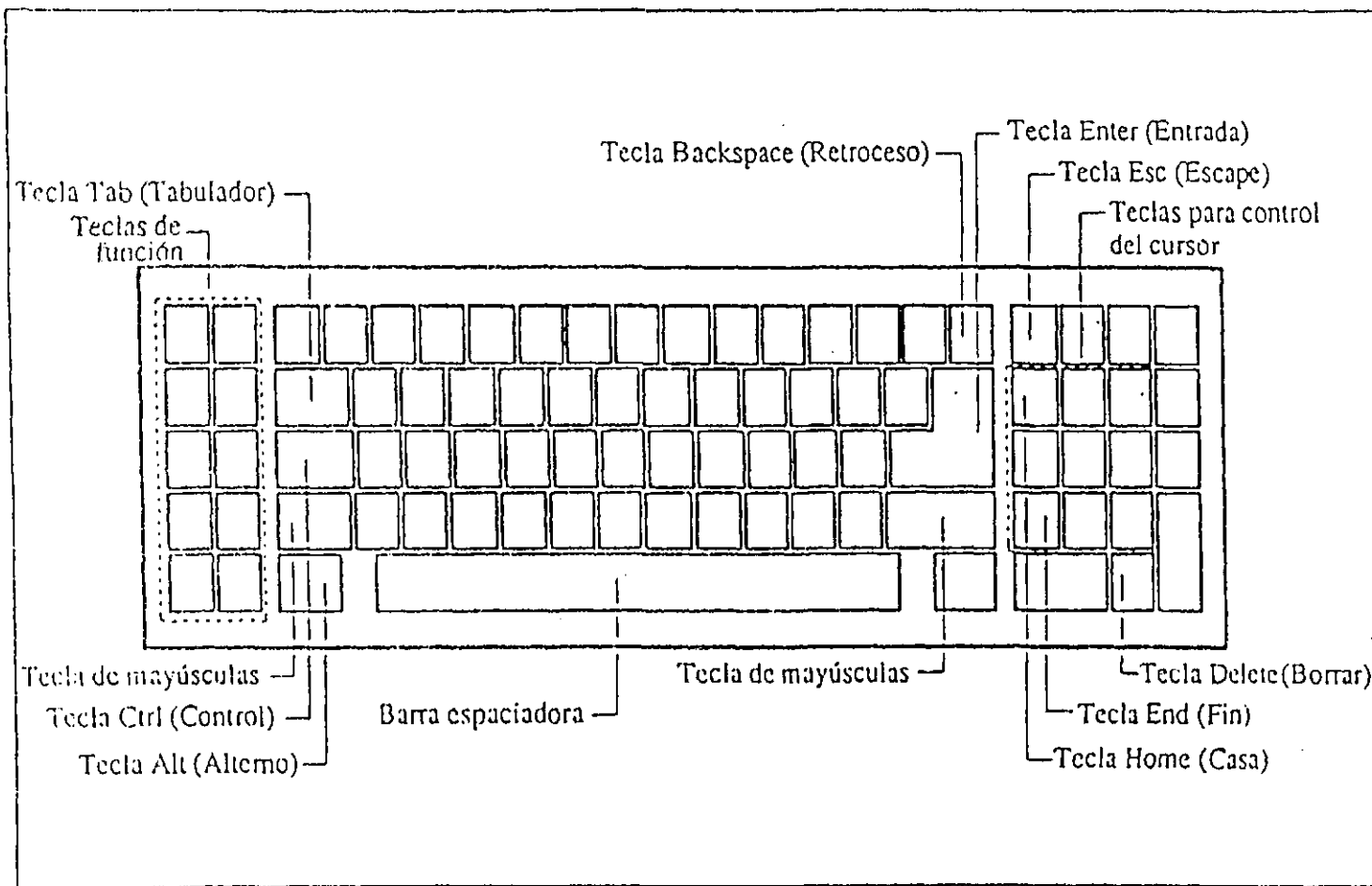


Figura 2-2. Teclado del IBM PC AT

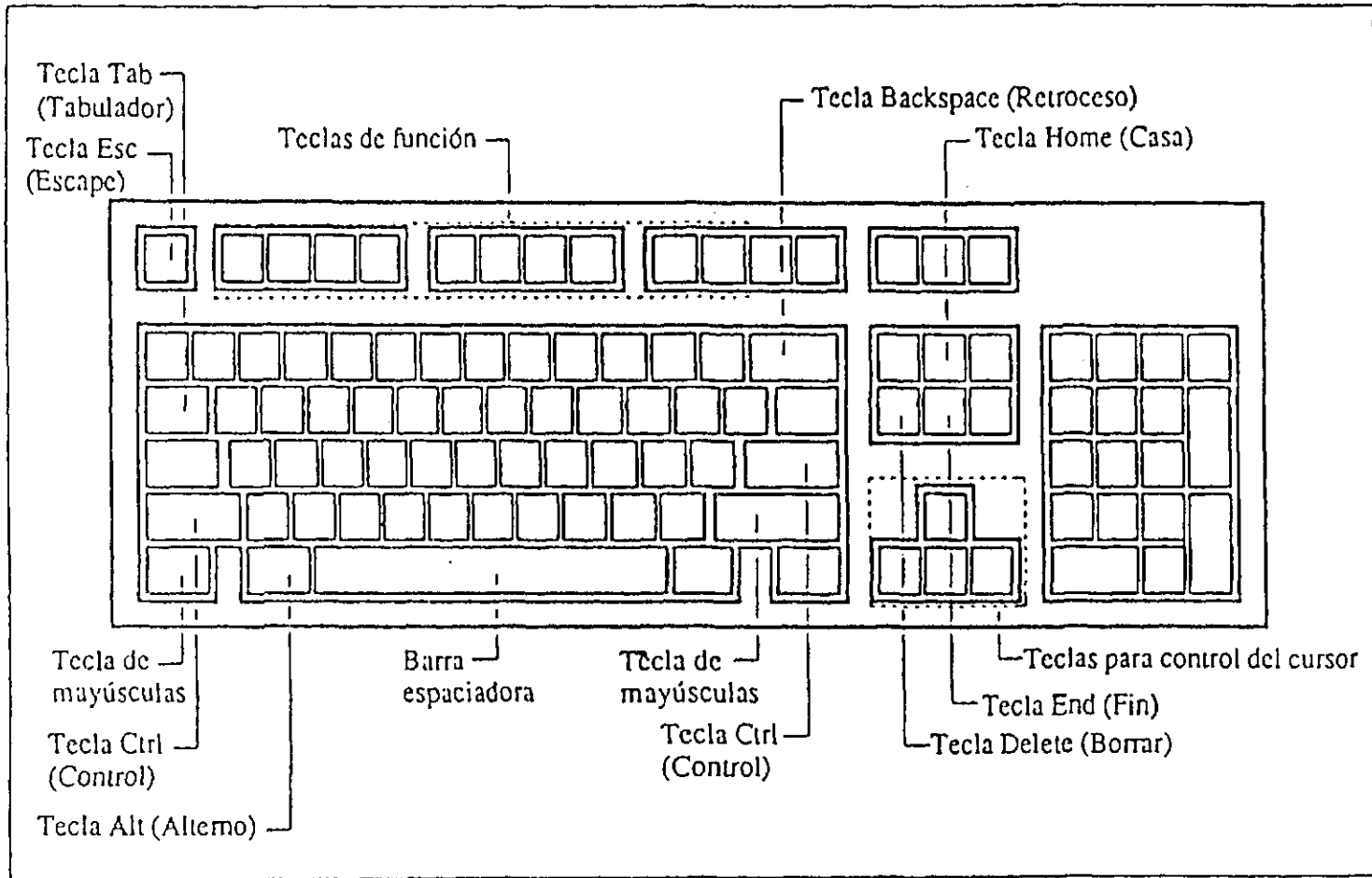


Figura 2-3. Teclado de 101 teclas

En esta guía (y, sobre todo, en este capítulo), aprenderás lo que hacen estas teclas importantes. Pero hasta que te hayas familiarizado con su posición, pon una marca en esta sección para que puedas volver a ella cuando te haga falta.

Ahora estás listo para poner en marcha STATGRAPHICS.

## **PUESTA EN MARCHA DE STATGRAPHICS.**

Hay tres pasos sencillos que dar para poner en marcha STATGRAPHICS. Pero, antes de darlos, tienes que saber el nombre del directorio (una zona en el disco del ordenador) donde se encuentran los ficheros de sistema de STATGRAPHICS. Si no estás familiarizado con el ordenador y con los mandos de DOS (Sistema Operativo de disco) usados en el ordenador, pídele a un amigo o a un compañero de trabajo que te ayude a determinar el nombre del directorio de STATGRAPHICS y a poner en marcha STATGRAPHICS.

- 1.- De ser necesario, enciende el ordenador y todo el equipo que tengas conectado a él (la impresora, por ejemplo).
- 2.- Cuando el ordenador esté listo para empezar a trabajar, verás en la pantalla un indicador de DOS (que generalmente es C: ). Tienes que ir a la zona del disco (el directorio) donde se encuentran los ficheros de sistema de STATGRAPHICS. Para ir a un directorio, se escribe CD (cambio de directorio) seguido de la barra inclinada invertida y del nombre del directorio, y luego se pulsa la tecla Enter. ( La tecla Enter se usa, para indicarle al ordenador que ya se ha terminado de escribir la información). Por ejemplo: para ir a un directorio llamado STATG, escribe CD STATG al aparecer el indicador de DOS.

```
C : CD STATG.
```

y luego pulsa la tecla Enter.

- 3.- Cuando el ordenador muestre el próximo indicador de DOS (C: STATG > en este ejemplo), escribe  
STATGRAF

```
C : STATG > STATGRAF
```

y pulsa la tecla Enter para poner en marcha STATGRAPHICS.

Después de esos tres pasos, en seguida STATGRAPHICS presenta en la pantalla varias líneas explicativas (la primera línea, "Loading STATGRAPHICS...", indica que se está cargando el programa).

Luego, STATGRAPHICS presenta una serie de mensajes en la pantalla. Primero se ve el "indicador de derechos de propiedad" con detalles sobre el sistema de STATGRAPHICS en uso, tal como se muestra en la Figura 2-4 (Si tu pantalla puede mostrar colores (si tienes un monitor de color), el indicador de derechos de propiedad empezará con un color y cambiará a varios colores después de unos momentos).

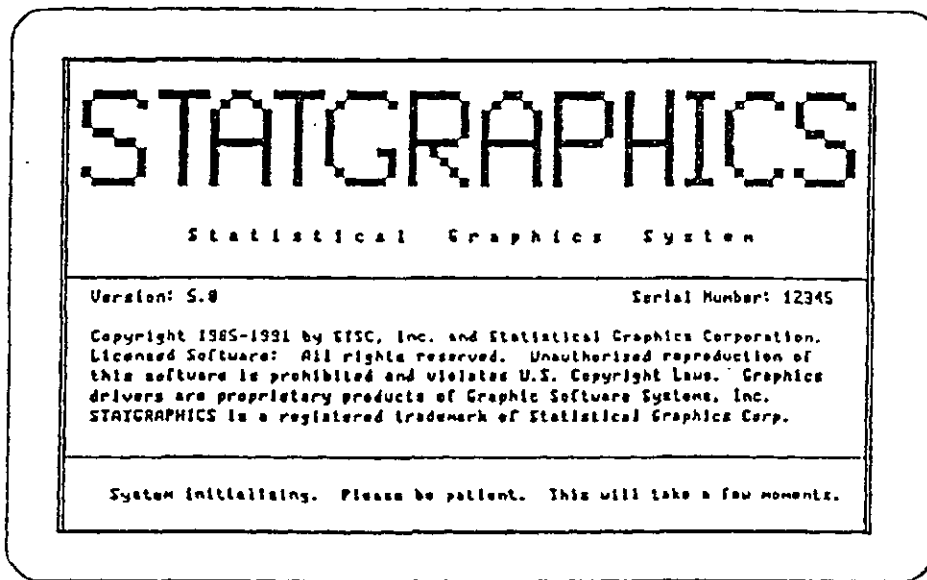


Figura 2-4. Indicador de derechos de propiedad de STATGRAPHICS

Luego se ve un mensaje, como en la Figura 2-5, diciendo que STATGRAPHICS está leyendo el directorio de datos. Esto significa solamente que STATGRAPHICS está preparando unos datos que hemos incluido para usarlos durante las prácticas con STATGRAPHICS.

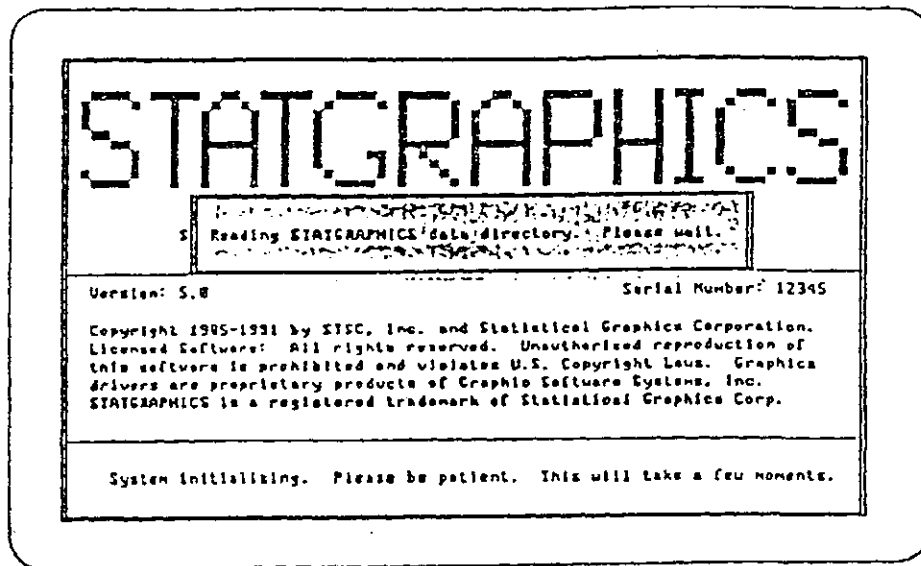


Figura 2-5. Mensaje de Lectura de datos del directorio de STATGRAPHICS

Por fin, aparece la pantalla (presentación de información que llena la pantalla del ordenador) llamada Menú principal, que se muestra en la Figura 2-6.

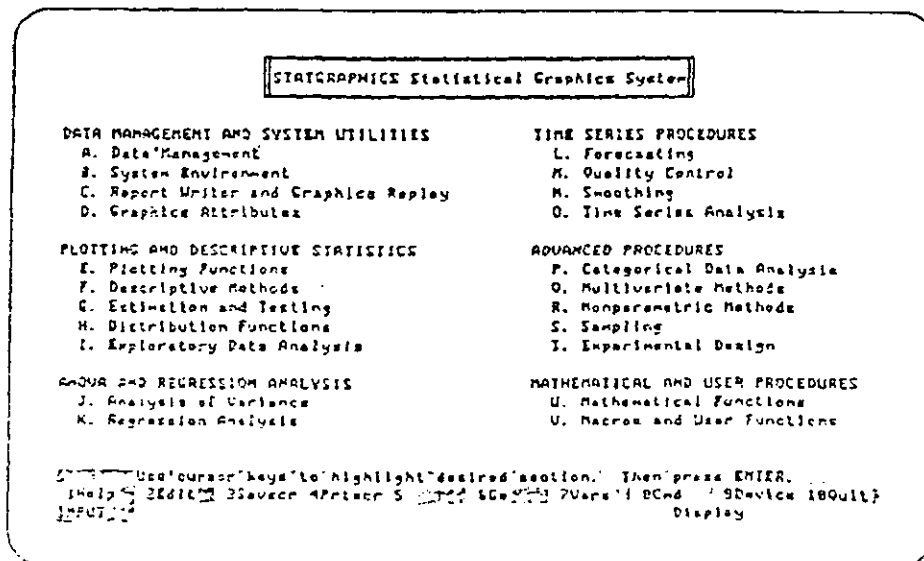


Figura 2-6. Menú principal de STATGRAPHICS

Como verás en seguida, el Menú principal facilita el acceso a los instrumentos estadísticos ofrecidos por STATGRAPHICS. Pero antes tienes que enterarte de una zona muy especial en las pantallas de STATGRAPHICS - la línea de estado.

## LA LINEA DE ESTADO.

Mira las tres últimas líneas del Menú principal que se muestra en la Figura 2-6. Estas tres líneas, juntas, forman la línea de estado. Excepto en los casos en que se esté mostrando una gráfica, siempre verás la línea de estado al final de la pantalla.

Aunque la línea de estado da mucha información, hay tres zonas que debes observar de cerca en esta guía : la primera línea, la del medio y los extremos de la última línea.

STATGRAPHICS usa la primera línea para mostrar instrucciones, información, y mensajes sobre errores. Por ejemplo : la primera línea de la línea de estado en la Figura 2-6 dice :

"Usa las teclas del cursor para seleccionar la sección deseada".  
Luego pulsa ENTER.

En este caso STATGRAPHICS dice cómo usar el Menú principal, sobre lo que aprenderás más en la sección siguiente.

STATGRAPHICS usa la última línea para informar sobre siete aspectos diferentes. Pero hay dos zonas especiales que debes vigilar durante tu trabajo con la guía : los extremos derecho e izquierdo de la última línea.

Como se ve en la Figura 2-6 , el extremo izquierdo de la última línea dice :

INPUT

STATGRAPHICS usa esta zona de la línea de estado para decirte lo que está haciendo el sistema. Cuando STATGRAPHICS muestra la palabra INPUT en esta zona, el sistema está esperando a que tú le digas lo que quieres hacer. Cuando STATGRAPHICS muestra la palabra PROCESS en esta zona, el sistema está haciendo lo que tú le hayas dicho que haga. Conforme vayas por la guía verás que STATGRAPHICS muestra, por lo menos, otra palabra en esta zona. Fijate con cuidado a ver si puedes determinar lo que STATGRAPHICS te diga.

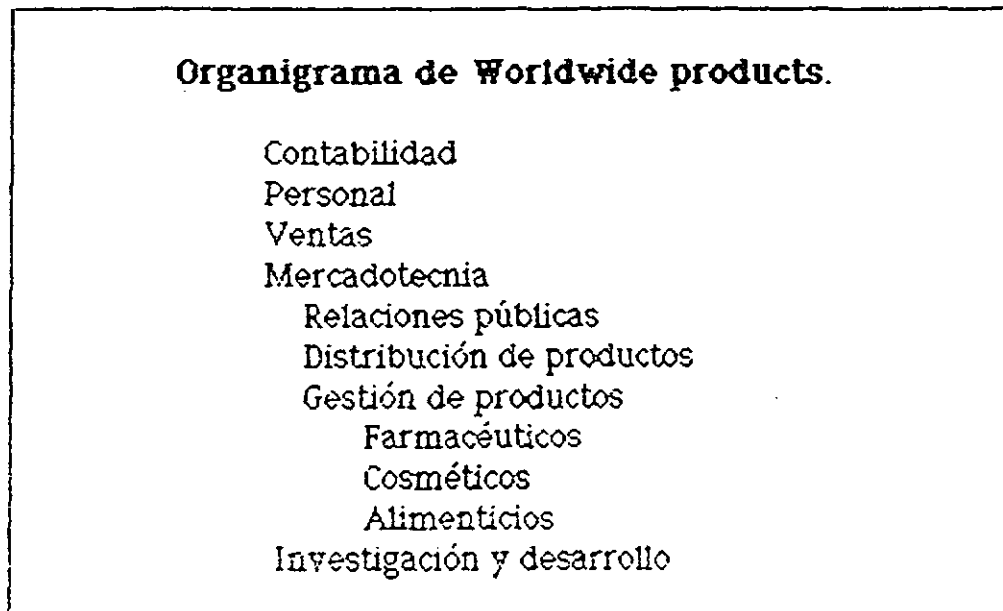
La segunda zona especial, en el extremo derecho de la línea final, no aparece en la Figura 2-6. ¿Por qué? Porque esta zona solamente se muestra después de que se

haya seleccionado un instrumento estadístico con el que trabajar. Antes de que termines de leer este capítulo verás aparecer la palabra BAR en esta zona. Mira a ver si la ves y si puedes determinar lo que STATGRAPHICS te quiere decir.

Ahora estás listo para enterarte de los procedimientos estadísticos que ofrece STATGRAPHICS, y para aprender el sistema de menús que usa STATGRAPHICS para organizarlos.

## LOS MENUS.

La organización del sistema de menús de STATGRAPHICS es semejante al organigrama de corporación de Worldwide Products (porciones del cual se muestran en la Figura 2-7)





Los empleados de Worldwide Products están agrupados en divisiones como contabilidad, personal, ventas, mercadotecnia, e investigación y desarrollo. Cada división de la compañía tiene varias secciones. Por ejemplo : dentro de la división de mercadotecnia hay secciones separadas para relaciones públicas, distribución de productos y gestión de productos. Dentro de la sección de gestión de productos hay tres gerentes para tres tipos diferentes de productos - farmacéuticos, cosméticos y alimenticios.

Cuando se pone en marcha STATGRAPHICS, se ve el Menú principal (véase la figura 2-8). En forma parecida a como el organigrama usa las divisiones de contabilidad, personal, ventas, mercadotecnia, e investigación y desarrollo para agrupar a los empleados de acuerdo con la función principal de su trabajo, el Menú principal de STATGRAPHICS usa seis categorías para agrupar los medios estadísticos de acuerdo con su función principal:

- Gestión de datos y Utilitarios de sistema
- Trazado y Estadística descriptiva
- ANOVA y Análisis de regresión
- Procedimientos de series temporales
- Procedimientos avanzados
- Procedimientos matemáticos y del usuario.

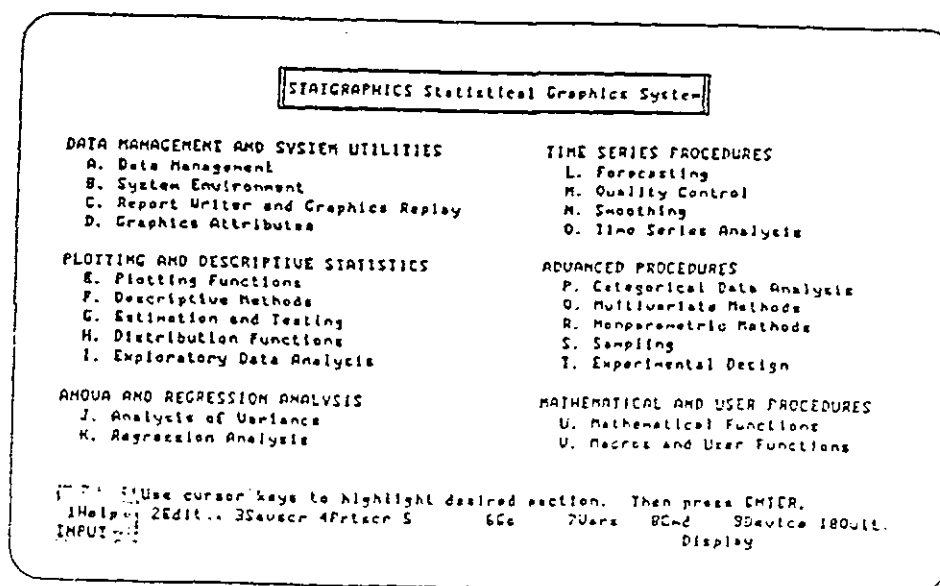


Figura 2-8. Menú principal de STATGRAPHICS

El Menú principal de STATGRAPHICS también hace referencia a 22 categorías secundarias dentro de las seis categorías principales (véase la figura 2-8). Las categorías secundarias como Gestión de datos, Entorno del sistema, y Escritor de informes y Visualizador de gráficas (en la categoría de Gestión de datos y Utilitarios del sistema) del Menú principal de STATGRAPHICS son similares a las secciones de relaciones públicas, distribución de productos, y gestión de productos (en la División de mercadotecnia) del organigrama.

Hasta aquí, el Menú principal de STATGRAPHICS es igual que el organigrama de Worldwide Products. Pero en el organigrama hay tres categorías secundarias productos farmacéuticos, cosméticos y alimenticios que figuran como funciones de la gestión de productos. ¿Dónde figuran los instrumentos estadísticos en el Menú principal de STATGRAPHICS? Por ejemplo : ¿Dónde están los instrumentos estadísticos para la categoría secundaria de Gestión de datos?

A diferencia del organigrama, STATGRAPHICS no puede mostrar de una vez todos los medios estadísticos de que dispone. Por ello, cada categoría secundaria del Menú principal de STATGRAPHICS tiene otro menú asociado con ella un menú secundario. Cada menú secundario indica los instrumentos estadísticos que hay disponibles en esa categoría secundaria.

La sección siguiente de esta guía explica cómo ver los menús secundarios en la pantalla. Pero, por ahora, examina la Figura 2-9 que muestra el menú secundario para la categoría secundaria de Gestión de datos.

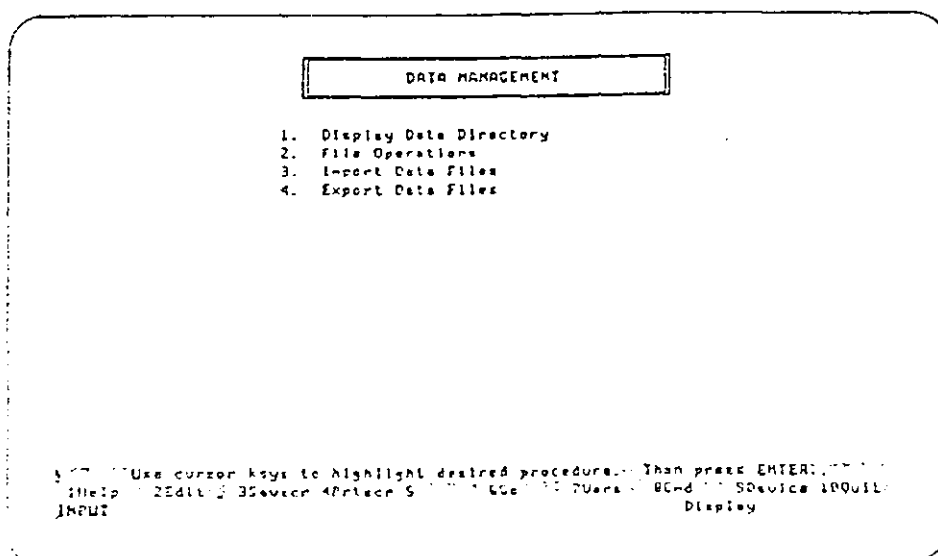


Figura 2-9. Menú secundario de gestión de datos

El menú secundario de Gestión de datos que aparece en la Figura 2-9 indica los procedimientos estadísticos usados para meter y manejar los datos — el procedimiento de Operaciones con ficheros, el procedimiento de Importación de ficheros de datos, etc. Observa que el título del menú secundario (enmarcado en la parte superior de la pantalla) es también el nombre de la categoría secundaria — Gestión de datos. Por ello, en el resto de la guía, la expresión menú secundario hace referencia a una categoría secundaria (como Gestión de datos o Funciones de trazado) del Menú principal. Además, en el resto de la guía el término procedimiento se refiere a cualquiera de los instrumentos estadísticos (como Operaciones con ficheros e Importación de ficheros de datos) ofrecidos por STATGRAPHICS.

Ahora que sabes lo que es el Menú principal y lo que son los menús secundarios, estás listo para aprender como llegar a un menú secundario y seleccionar un procedimiento estadístico.

## **SELECCION DE PROCEDIMIENTOS.**

Imagínate que la sección de personal quiere que les hagas una gráfica de barra. (La gráfica de barra, una gráfica bastante común, usa barras horizontales o verticales para representar cantidades). La gráfica de barra tendrá que mostrar el número de empleados en cada una de las cinco divisiones de Worldwide Products: contabilidad, personal, ventas, mercadotecnia, e investigación y desarrollo. ¿Por dónde empezar?

Cuando pones en marcha STATGRAPHICS y el sistema presenta el Menú principal, el cursor (el subrayado intermitente que indica en qué parte de la pantalla estás) se encuentra en el menú secundario de Gestión de datos (véase la Figura 2-10).

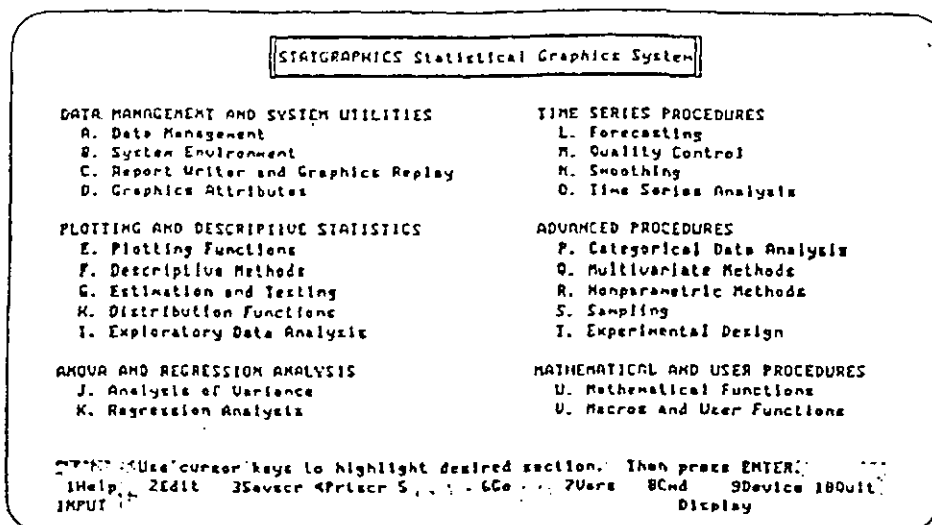


Figura 2-10. El menú principal

Nota también que "Data Management" (Gestión de datos) aparece enmarcado en color. Cuando una parte del Menú principal o de un menú secundario aparece enmarcada en color, decimos que está "seleccionada". (Si tienes monitor de color-si la pantalla del ordenador puede mostrar muchos colores-el enmarcado puede ser del color que elijas (o puedes dejar el color que STATGRAPHICS lo seleccionen). Si la pantalla puede mostrar un color solamente (se trata de un monitor monocromo), el enmarcado siempre aparece en ese color).

El primer paso para seleccionar un procedimiento estadístico es ir al menú en el que esté el procedimiento. Para seleccionar un menú secundario desde el Menú principal, usa las teclas que controlan el movimiento del cursor. A estas teclas las llamamos teclas de control del cursor, y están situadas en la parte derecha del teclado (de ser necesario, consulta de nuevo los diagramas de teclado en las Figuras 2-1, 2-2, 2-3).

Antes de seguir adelante, prueba a pulsar la tecla de control del cursor que tiene la flecha hacia la derecha. El cursor (y la selección) se van a la derecha, al menú secundario de Pronóstico. Pulsa la tecla con la Flecha hacia abajo. Se selecciona el menú secundario de Control de calidad. Prueba con la tecla End (fin). El seleccionado ahora es el último menú secundario-Macros y Funciones del usuario. Pulsa la tecla con la Flecha hacia la izquierda. Se selecciona el menú secundario de Análisis de regresión. Pulsa la tecla con la Flecha hacia arriba. Se selecciona el

menú secundario de Análisis de Varianza. Para terminar, pulsa la tecla Home (casa). La selección vuelve al primer menú secundario- Gestión de datos. Como ves, se usan las teclas de control del cursor para moverlo (y seleccionar) diferentes puntos de la pantalla.

Antes de volver a la tarea original, pulsa la tecla de función F1. STATGRAPHICS muestra una pantalla que contiene información para ayudarte (véase la Fig. 2-11). Ezepto en los casos en que se esté viendo una gráfica, la tecla de función F1 (Help) (Ayuda) está siempre a tu disposición - para recordarte cómo llegar a un procedimiento estadístico o para describir uno que quieras emplear. Mira ahora la línea de estado. La primera línea te dice "Press Esc to return" (Pulsa Esc para volver). Pulsa la tecla Esc (Escape). El sistema vuelve a la pantalla de STATGRAPHICS en la que estabas - el Menú principal.

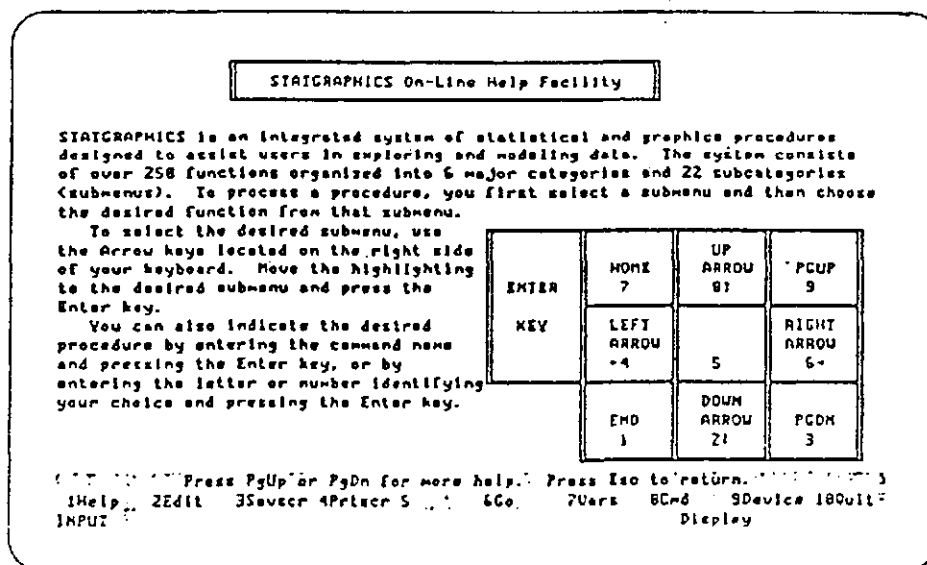


Figura 2-11. Servicio de ayuda en línea de STATGRAPHICS

Ahora estás listo para volver a la tarea original- producir una gráfica de barra de los empleados en cinco divisiones de Worldwide Products. El menú secundario de Funciones de trazado contiene el procedimiento de Gráficas de barra. Para producir una gráfica de barra tienes que ir a menú secundario de Funciones de trazado. Pulsa la flecha hacia abajo (control del cursor) hasta que quede seleccionado el menú secundario de Funciones de trazado (véase la Figura 2-12).

(Si pulsas la Flecha hacia abajo más veces de las necesarias, usa la Flecha hacia arriba para corregir el error).

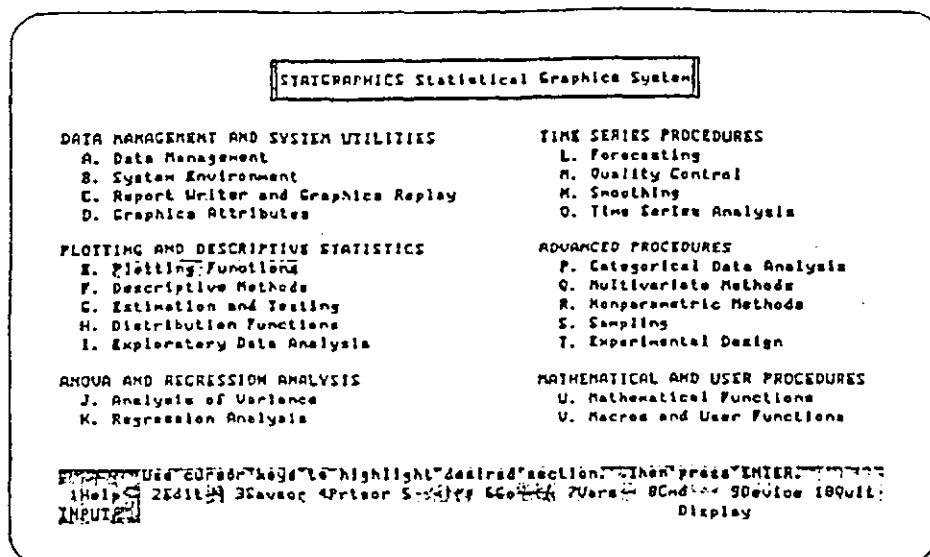


Figura 2-12. Menú principal con selección del Menú secundario de Funciones de trazado

A continuación tienes que ver los procedimientos (instrumentos estadísticos) disponibles en el menú secundario de Funciones de trazado. Una vez seleccionado este menú secundario de Funciones de trazado, pulsa la tecla Enter. (Se usa la tecla Enter para decirle a STATGRAPHICS que vaya al menú secundario seleccionado para procesar lo que se selecciona en el menú secundario). STATGRAPHICS presenta en la pantalla el menú secundario de Funciones de trazado en la pantalla (véase la Figura 2-13). Este menú secundario indica que hay siete procedimientos (instrumentos estadísticos) uno de los cuales es Gráficas de barra.

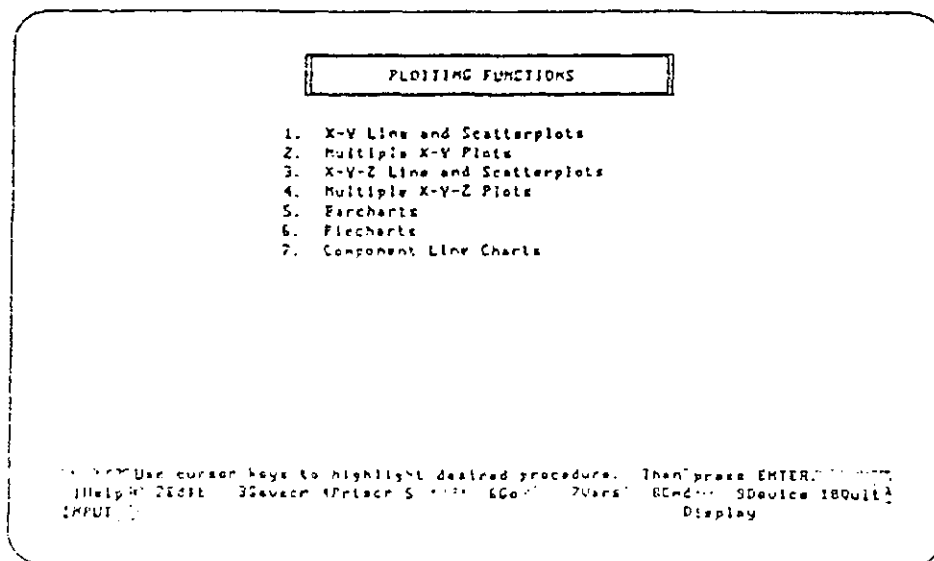


Figura 2-13. Menú secundario de Funciones de trazado

Desde un menú secundario se llega a un procedimiento de la misma forma que se llega a un menú secundario desde el Menú principal- empleando las teclas de control del cursor. Pulsa la Flecha hacia abajo varias veces hasta que quede seleccionado el procedimiento de Gráficas de barra. Luego pulsa la tecla Enter para activar la selección en el menú secundario y llegar al procedimiento de Gráficas de barra. STATGRAPHICS muestra la pantalla de entrada de datos para la Gráfica de barras (véase la Figura 2-14 ) la primera de varias pantallas asociadas con el proceso de Gráficas de barra. En la próxima sección, aprenderás lo relacionado con las pantallas de entrada de datos.

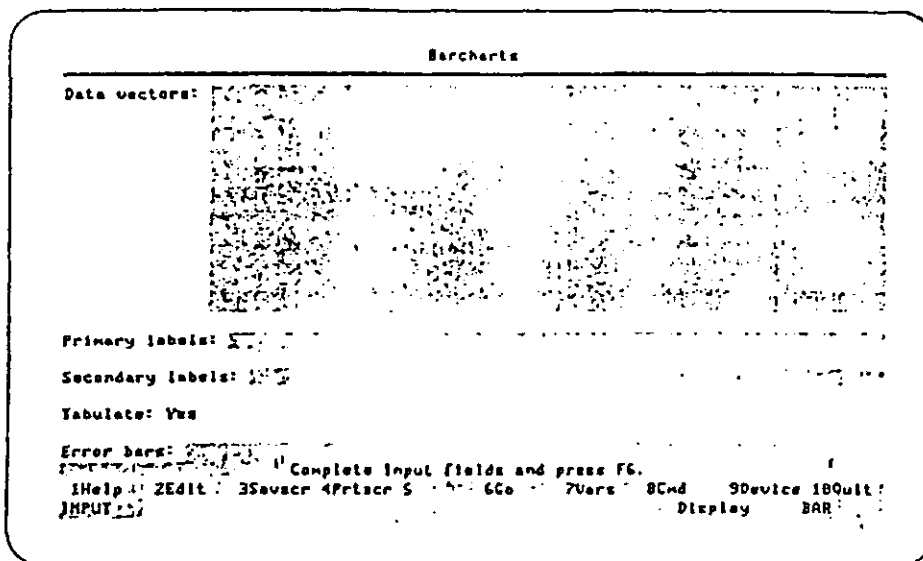


Figura 2-14. Pantalla de entrada de datos para Gráficas de barra

Nota: En esta guía se usa el Menú principal y los secundarios para llegar a los procedimientos estadísticos. Pero STATGRAPHICS permite emplear mandos (como BAR para el procedimiento de Gráficas de barra) para llegar a los procedimientos estadísticos. Una vez te familiarices más con los procesos de STATGRAPHICS, podrás usar mandos para llegar a los procedimientos de manera más directa que a través de los menús.

Observa que cuando llegas a un procedimiento estadístico (como el procedimiento de Gráficas de barra), STATGRAPHICS muestra el nombre del mando asociado en la línea de estado en el extremo derecho de la tercera línea (véase la Figura 2-14 ).

## PANTALLA DE ENTRADA DE DATOS.

Mira la pantalla de entrada de datos de Gráficas de barra en la Figura 2-14. Observa que :

- la palabra "Barcharts" (Gráficas de barra) aparece en la parte superior de la pantalla.
- hay cinco grupos de palabras a la izquierda de la pantalla (Data Vectors, Primary Label, etc.).
- cada grupo de palabras tiene un cuadro a la derecha.
- si la pantalla es de color, un cuadrado (el que está al lado de las palabras "Data Vectors") aparece destacada de un color diferente del que tienen los otros.
- un cuadrado (el que está junto a la palabra "Tabulate") ya tiene información en él.

La palabra "Barcharts" (Gráficas de barra) en la parte superior de la pantalla indica que ésta es la pantalla para la entrada de datos para el procedimiento de Gráficas de barra. Las palabras a la izquierda de la pantalla se llaman nombres de campo, y los cuadrados adyacentes se llaman campos. Campos son zonas de la pantalla usadas para la entrada de datos o para decirle a STATGRAPHICS lo que se desea hacer. Los nombres de campo indican la información que se debe escribir en los campos correspondientes.

El cuadrado que aparece diferente en un monitor de color se llama el campo activo. El campo activo siempre es el campo en el que se encuentra el cursor. Por eso, conforme se mueve el cursor de campo a campo, cambia el campo activo. (Nota: Si la pantalla tiene un solo color, todos los campos son del mismo color, pero el campo activo es el que tiene el cursor).

Observa también que STATGRAPHICS ya ha puesto la palabra Yes (Sí en el campo Tabulate (Tabular).

Cualquier palabra o número puesto por STATGRAPHICS en un campo se llama opción por defecto. La opción por defecto que STATGRAPHICS ofrece es el valor más comúnmente usado en ese campo. Si tú no indicas un valor diferente en el campo, STATGRAPHICS usa la opción por defecto al procesar el procedimiento estadístico.



A veces, el campo que contiene la opción por defecto es también un campo conmutable uno en el que no se puede poner información, pero cuyo valor se puede cambiar. ¿Cómo se puede cambiar el valor si no se puede meter información?

En la próxima sección aprenderás cómo poner datos en un campo y cómo cambiar el valor del un campo conmutable.

## ENTRADA DE DATOS.

Ahora que ya has llegado a la pantalla de entrada de datos para Gráficas de barra (véase la Figura 2-15), estás listo para decirle a STATGRAPHICS el tipo de gráfica de barra que deseas producir. Se quiere crear una gráfica de barra sencilla de los empleados de Worldwide Products para mostrar el número de empleados en cada una de las cinco divisiones.

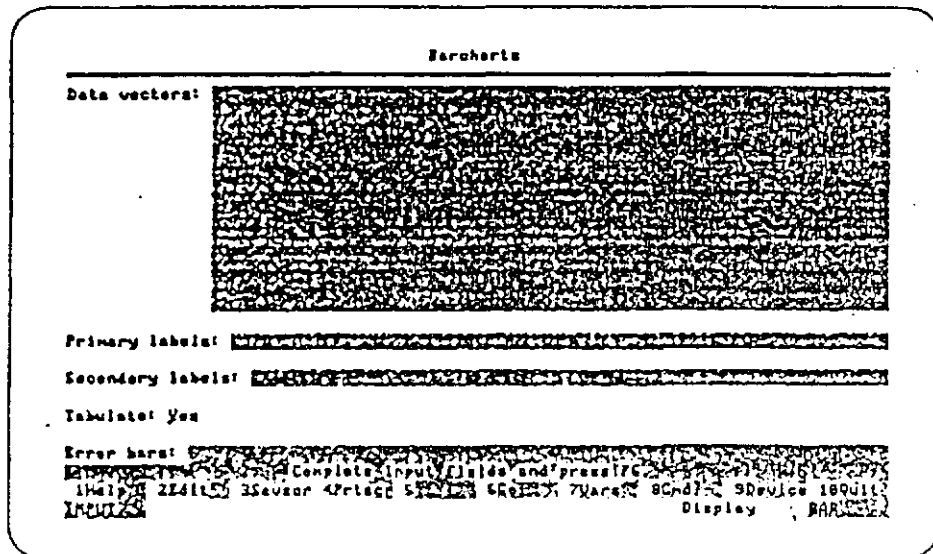


Figura 2-15. Pantalla de entrada de datos para Gráficas de barra

Primero, tienes que darle a STATGRAPHICS los datos que quieras analizar - información sobre las divisiones en las que trabajan los empleados. Para este ejemplo vamos a suponer que hay un total de 15 empleados y que están repartidos por las cinco divisiones en la forma siguiente:

- 1 empleado en la División 1 (Contabilidad)
- 2 empleados en la División 2 (Personal)
- 3 empleados en la División 3 (Ventas)
- 4 empleados en la División 4 (Mercadotecnia)
- 5 empleados en la División 5 (Investigación y desarrollo).

Cuando aparece por primera vez la pantalla de entrada de datos para Gráficas de barra, el cursor se encuentra en el campo Data Vectors (Vectores de datos); es decir, el campo Vectores de datos es el campo activo. Escribe los números siguientes en el campo de Vectores de datos (cerciórese de poner entre los números un espacio como delimitador):

1 2 2 3 3 3 4 4 4 4 5 5 5 5 5

Como verás, los números representan a los empleados según la división en la que trabajan un empleado en la división 1, dos empleados en la división 2, etc. Si te equivocas al escribir los números, puedes usar la tecla Backspace (Retrosceso) para borrar hacia atrás hasta llegar al error y poder escribir correctamente los números.

No hace falta usar etiquetas (nombres como Contabilidad, Personal, etc.) para las barras de un trazado tan sencillo, así que omite los campos Primary Labels y Secondary Labels (Etiquetas primarias y secundarias).

¿Cómo se va de campo a campo en la pantalla de entrada de datos? Pulsa la Flecha hacia abajo (control del cursor). El cursor baja una línea en el campo Data Vectors. Pulsa la tecla End. El cursor va al principio de la línea. Prueba la Flecha hacia la derecha y la Flecha hacia la izquierda. El cursor se desplaza una posición en la dirección correspondiente. Para pasar de un campo a otro en la pantalla de entrada de datos usa la tecla Tab.). Así pues, para pasar el cursor al campo Tabulate y activarlo, pulsa la tecla Tab tres veces.

Antes aprendiste que no se puede meter información en un campo conmutable, pero que se puede cambiar el valor que contiene. ¿Cómo es posible? En un campo conmutable, STATGRAPHICS pone todos los valores posibles. Pulsando la Barra espaciadora repetidas veces se pasa (conmuta) por el ciclo de todos los valores posibles. Después de activar el campo Tabulate (con el cursor dentro de él), pulsa la Barra espaciadora una vez. Ahora el campo Tabulate contiene la palabra No. Pulsa

la Barra espaciadora otra vez. El campo Tabulate tiene de nuevo la palabra Yes (sí). Es decir, STATGRAPHICS da los dos valores posibles para este campo-Yes y No- y tú cambias (conmutas) el valor en el campo pulsando la Barra espaciadora hasta llegar al valor deseado.

Antes de producir la Gráfica de barra, hay una cosa más que tienes que saber en relación con la entrada de información en las pantallas de STATGRAPHICS cómo se comunica STATGRAPHICS contigo por medio de señales auditivas. Lo aprenderás en la próxima sección.

## SEÑALES AUDITIVAS.

Si tu ordenador puede producir sonidos, STATGRAPHICS usa dos señales diferentes para comunicarse contigo. Las llamamos trino y bocina. Estas señales te dicen que has hecho un error.

En este momento, la pantalla del ordenador debería aparecer como se muestra en la Figura 2-16.

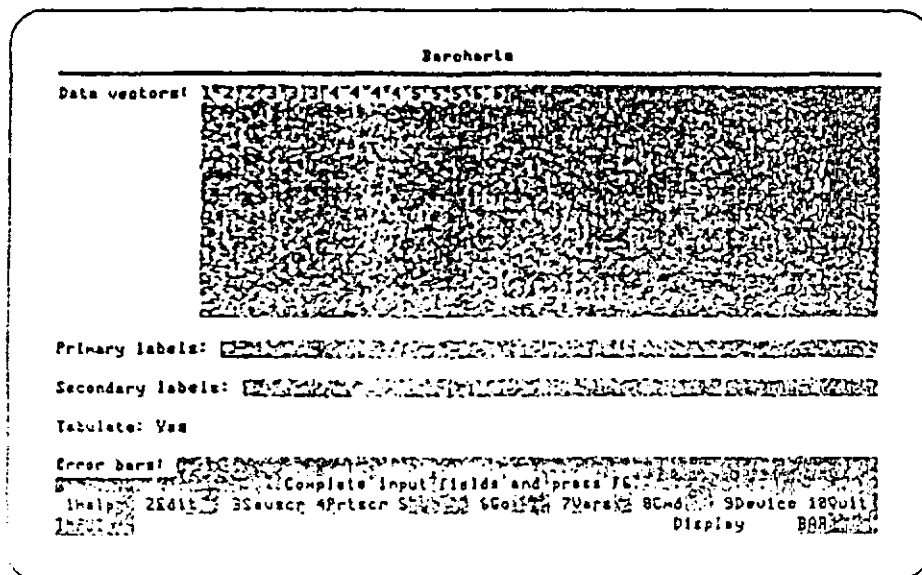


Figura 2-16. Pantalla de entrada de datos para Gráficas de barra con los datos de los empleados

Usa la tecla Tab para activar el campo Error Bars (Barras de error). Luego pulsa la tecla F5 varias veces. El sistema hará sonar un trino cada vez que pulses la tecla F5, para decirte que, en este momento, la tecla F5 no hace nada. (De hecho, el

sistema hará sonar un trino cada vez que pulses cualquier tecla que no sea adecuada para la acción a llevar a cabo).

Ahora escribe tu nombre de pila en el campo Error Bars y pulsa la tecla de función F6 (Go) (Adelante) para decirle a STATGRAPHICS que procese la información puesta en la pantalla. El sistema hace un sonido de bocina para indicarte que hay una entrada que no está bien. ¿Qué es lo que está mal? La primera línea de la línea de estado explica el error:

ERROR : cannot plot error bars when data is being tabulated. (No se puede trazar barras de error mientras se tabulan datos).

Siempre que oigas la bocina, mira la línea de estado; comprueba la información entrada e intenta de nuevo.

Antes de que se pueda procesar esta pantalla tienes que corregir el error-quita tu nombre del campo Error Bars. Usa la tecla Tab para activar el campo Error Bars, y luego borra tu nombre con la tecla Delete (Borrar), letra por letra. Ahora pulsa F6 (Go) (Adelante). ¿Qué pasa?

STATGRAPHICS muestra una pantalla nueva llamada Opciones de Gráficas de barra (Véase la Figura 2-17 ) y hace sonar la campana para decirte que el sistema te está esperando. Observa que se usa la tecla de función F6 (Go) (Adelante) para decirle a STATGRAPHICS que procese la información.

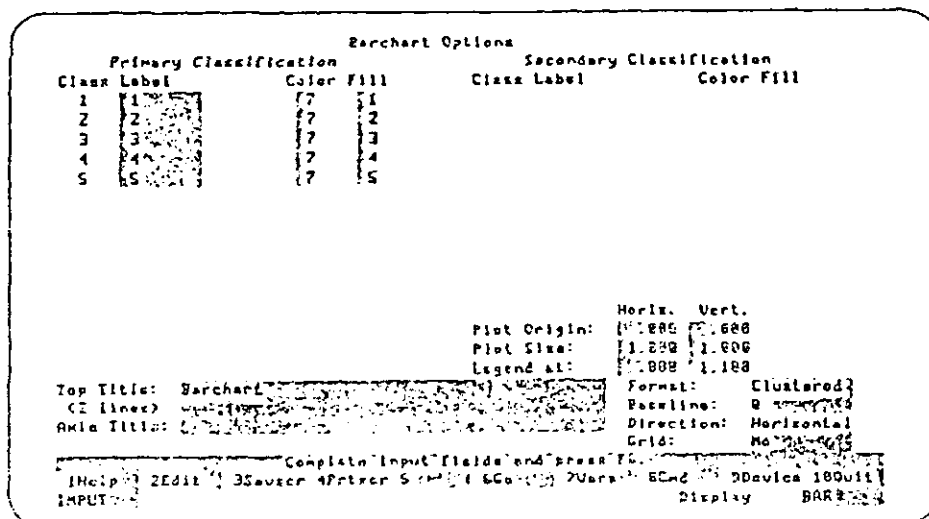


Figura 2-17. Pantalla de opciones para Gráficas de barra

Como se ve en la Figura 2-17, la pantalla de Opciones de Gráficas de barra muestra (y permite que se cambien) muchas opciones por defecto, incluso el título en la parte superior de la gráfica de barra. Vamos a usar las opciones por defecto en esta pantalla, así que estás listo para generar la gráfica de barra. Pulsa F6 (Go) (Adelante) y STATGRAPHICS mostrará en la pantalla la gráfica de barra de empleados por número de división (Véase la Figura 2-18).

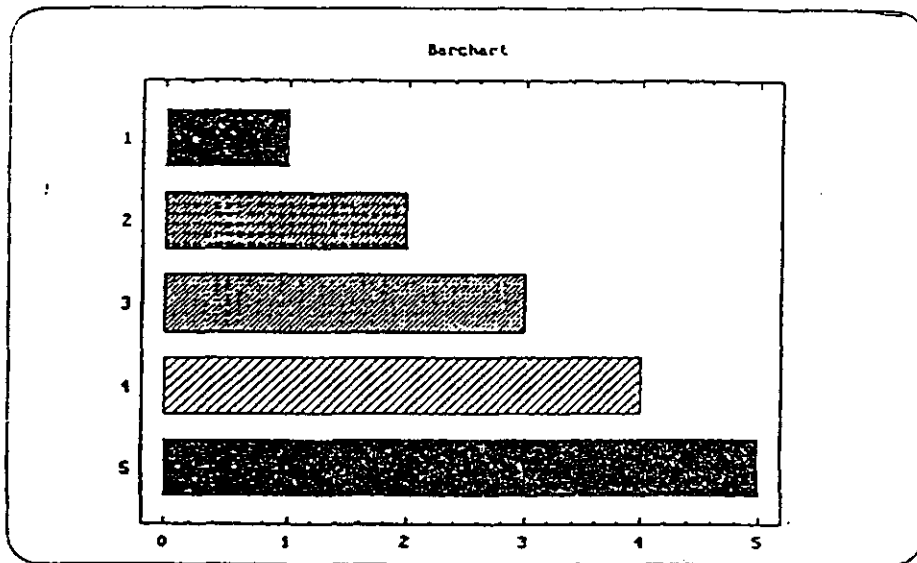


Figura 2-18. Gráfica de barra de los empleados por número de división

Ahora que ya has generado la gráfica de barra, estás listo para volver al Menú principal. Siempre que quieras terminar el proceso en marcha o salir de la pantalla en uso, pulsa la tecla Esc (Escape). Para volver al Menú principal, pulsa la tecla Esc varias veces hasta que aparezca el Menú principal en la pantalla. (Si pulsas la tecla Esc desde el Menú principal, STATGRAPHICS te preguntará si quieres terminar la sesión (salir de STATGRAPHICS). Para corregir el error y volver al Menú principal, pulsa la letra N (No) en el teclado).

¡Enhorabuena! Has aprendido mucho sobre cómo usar STATGRAPHICS. Pero hay unas cuantas cosas más que tienes que aprender antes de pasar al Capítulo 3. Tienes que entender como usar títulos y variables para almacenar los datos de STATGRAPHICS, y tienes que aprender los pasos apropiados para salir de STATGRAPHICS.

## FICHEROS Y VARIABLES.

Cuando se empieza un proyecto, lo primero es reunir los materiales necesarios. Por ejemplo, si durante el fin de semana planeas archivar los papeles que se hayan amontonado en una esquina de la mesa del despacho, es probable que durante la semana vayas a buscar carpetas, etiquetas y bolígrafos de colores y que los guardes en un cajón hasta el fin de semana cuando estés listo para empezar a archivar.

Antes de usar un procedimiento de STATGRAPHICS, tienes que reunir todos los materiales necesarios. En el caso de los procedimientos de STATGRAPHICS, generalmente los materiales son variables que contienen juegos de valores de datos relacionados. Por ejemplo, acaso quieras emplear una variable llamada **complaints** (quejas) para tomar nota de comentarios negativos recibidos sobre las hamburguesas de Worldwide Products. También podrías usar una variable llamada **hamburgers** para tomar nota del peso de las hamburguesas conforme salen de la línea de producción. Como en el caso de las etiquetas en los archivos, se puede modificar (editar) las variables siempre que no se las esté usando en un procedimiento de STATGRAPHICS.

De la misma manera que usas un cajón para guardar las carpetas y el resto de los materiales hasta el fin de semana, en STATGRAPHICS se usa un fichero para almacenar una o más variables relacionadas. Los ficheros permiten encontrar fácilmente las variables a la hora de usarlas. Por ejemplo, puedes emplear un fichero llamado QUALITY (CALIDAD) para agrupar las variables **complaints** y **hamburgers** (quejas y hamburguesas), puesto que las dos se relacionan con la calidad de fabricación de las hamburguesas de Worldwide Products.

Para comprender más fácilmente el concepto de ficheros y variables de STATGRAPHICS, piensa en un archivador de oficina. En el archivador empleas carpetas para guardar documentos relacionados. Por ejemplo, un archivador llamado CORPORACION para guardar comunicaciones distribuidas a toda la compañía y listas de números de teléfono de los empleados. O una carpeta en el archivador llamada PERSONAL para guardar listas detalladas de las horas que hayas trabajado y de las tareas que hayas llevado a cabo. Las carpetas se llaman ficheros en la terminología de STATGRAPHICS, y las hojas con información dentro de las carpetas se llaman variables.

En tu trabajo, usarás muchos ficheros y variables con información sobre diferentes aspectos del funcionamiento de Worldwide Products. En los Capítulos 3 y 4, aprenderás cómo crear ficheros y variables. En el Capítulo 5, aprenderás cómo tomar ficheros y variables de otras fuentes y usarlos con STATGRAPHICS. En otras

partes de la guía usarás ficheros y variables que te hemos proporcionado con STATGRAPHICS. Si instalaste los juegos de datos muestra al instalar STATGRAPHICS en el ordenador, esos ficheros y variables están preparados para que los uses.

Lo último que tienes que aprender es cómo salir de STATGRAPHICS. La próxima sección describe los pasos que tienes que dar para hacerlo.

### **SALIDA DE STATGRAPHICS.**

Hay tres pasos que tienes que dar siempre que quieras salir de STATGRAPHICS. (Si quieres continuar ahora con el Capítulo 3, deja esta sección hasta que estés listo para salir de STATGRAPHICS).

**PRECAUCION :** Para proteger programas y datos, sigue siempre estos tres pasos antes de apagar el computador.

- 1.- Pulsa la tecla Esc varias veces hasta que la pantalla muestre el Menú principal de STATGRAPHICS.
- 2.- Pulsa otra vez la tecla Esc. STATGRAPHICS pregunta con un mensaje en la pantalla: "si se desea poner termino a la sesión" con Y/N.
- 3.- Escribe una "Y" pulsa la tecla Enter.

### **RESUMEN.**

En este capítulo aprendiste :

- cómo poner en marcha STATGRAPHICS.
- cómo están organizados los procedimientos (instrumentos estadísticos) en STATGRAPHICS.
- cómo llegar a los muchos procedimientos ofrecidos por STATGRAPHICS.
- como se comunica STATGRAPHICS contigo.
- cómo usar variables y ficheros para almacenar los datos de STATGRAPHICS.
- cómo salir de STATGRAPHICS.

## MANEJO DE DATOS CON EL EDITOR DE DATOS.

En el Capítulo 2 aprendiste que hay dos formas de proporcionarle datos a STATGRAPHICS : creando ficheros y variables y usando ficheros y variables que provienen de otras fuentes. En este capítulo y en el Capítulo 4 vas a aprender cómo crear y modificar (editar) tus ficheros y variables que alguien te dé.

Por ejemplo, en las tres lecciones siguientes vas a crear variables con datos para 14 productos cosméticos fabricados por Worldwide Products. El fichero llamado COSMETIC tendrá las variables **sales**, **units**, **onhand**, y **products** (ventas, unidades, disponibles, y productos).

En STATGRAPHICS, las variables se crean, almacenan y editan empleando diferentes métodos. En este capítulo vas a aprender a trabajar con las variables empleando el editor de datos en el procedimiento de Operaciones con ficheros. (El editor de datos es un editor tipo "hoja de cálculo" que permite editar (modificar) una tabla de información).

## CREACION DE VARIABLES NUMERICAS.

En esta lección crearás un fichero nuevo llamado COSMETIC empleando el procedimiento de Operaciones con ficheros. También crearás y almacenarás tres variables numéricas- **sales**, **units**, y **onhand**- dentro del fichero COSMETIC empleando el editor de datos del procedimiento de Operaciones con fichero.

La Lección empieza en el Menú principal.

A la hora de crear ficheros y variables nuevos, el primer paso es usar los menús para seleccionar el procedimiento de Operaciones con ficheros. El procedimiento de Operaciones con ficheros está en el menú secundario de Gestión de datos, así que empieza seleccionando el menú secundario de Gestión de datos en el Menú principal (véase la Figura 3-1).

- De ser necesario, usa la tecla Home (control del cursor), para seleccionar el menú secundario de Gestión de datos.



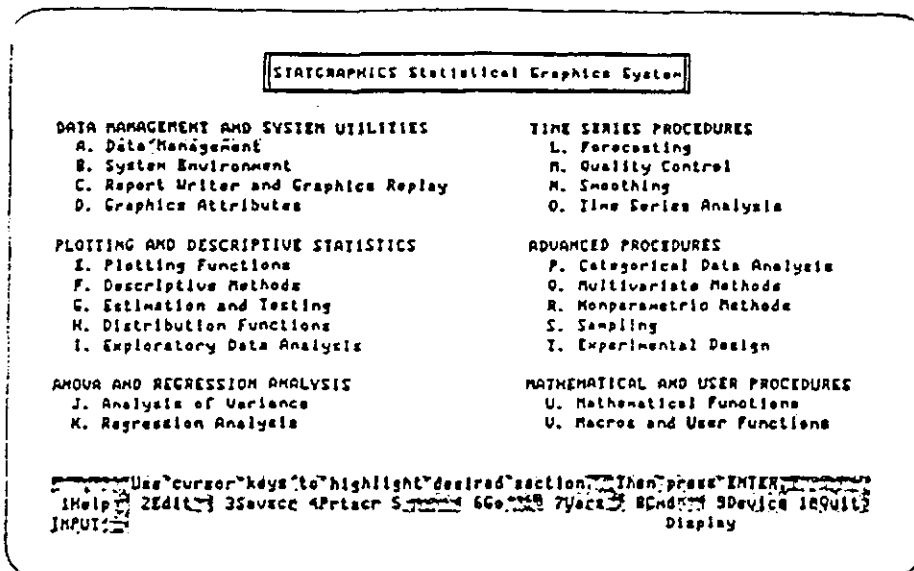


Figura 3-1. Menú principal con el menú secundario de Gestión de datos seleccionado

- Pulsa la tecla Enter.

Una vez se pulsa la tecla Enter para procesar la selección en el Menú principal, el sistema presenta el menú secundario de Gestión de datos (véase la Figura 3-2).

Ahora puedes seleccionar el procedimiento de Operaciones con ficheros en el menú secundario de Gestión de datos.

- Usa la Flecha hacia abajo (control del cursor) para seleccionar el procedimiento de Operaciones con ficheros.

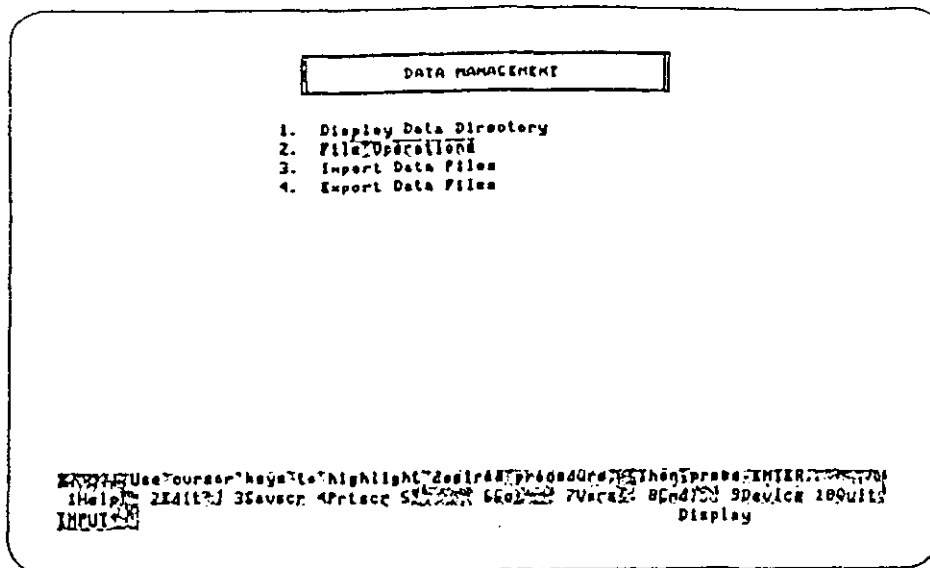


Figura 3-2. Menú secundario de Gestión de datos con el procedimiento de Operaciones con ficheros seleccionado

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar la selección en el menú secundario de Gestión de datos, el sistema presenta la pantalla de entrada de datos para Operaciones con ficheros (véase la Figura 3-3). Se usa esta pantalla para indicar el fichero que se desea usar y la operación a llevar a cabo.

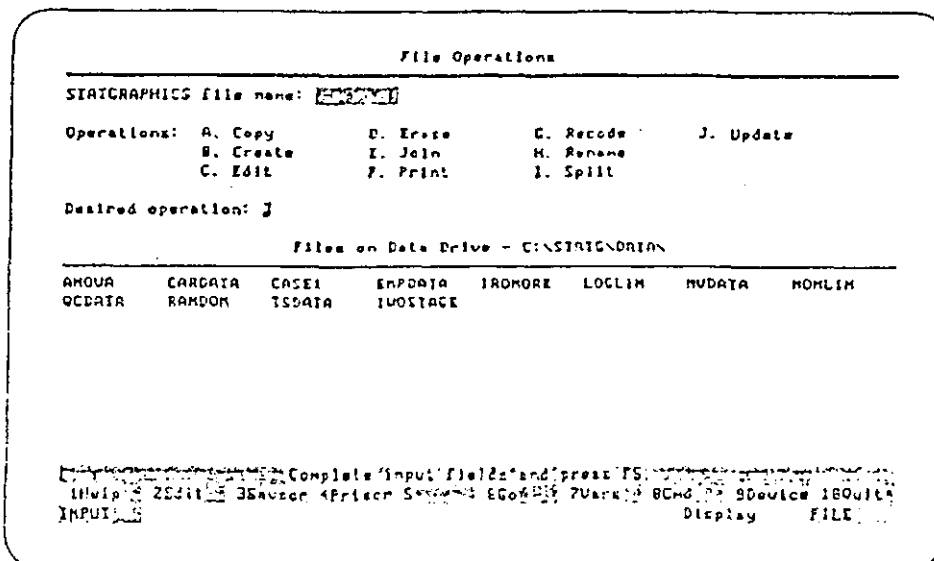


Figura 3-3. Pantalla de entrada de datos para Gestión de ficheros

Ahora puedes indicar el nombre del fichero y la operación a llevar a cabo. En esta lección se va a emplear la operación de Editar para crear un fichero nuevo llamado COSMETIC y entrar en el editor de datos. Esta información se especifica en la pantalla de entrada de datos para Operaciones con ficheros (véase la Figura 3-4).

- Escribe el nombre del fichero cosmetic en el campo File Name de STATGRAPHICS. El cursor se desplaza automáticamente al campo Desired Operation (Operación deseada).
- Escribe la letra C (para la operación Edit) en este campo.

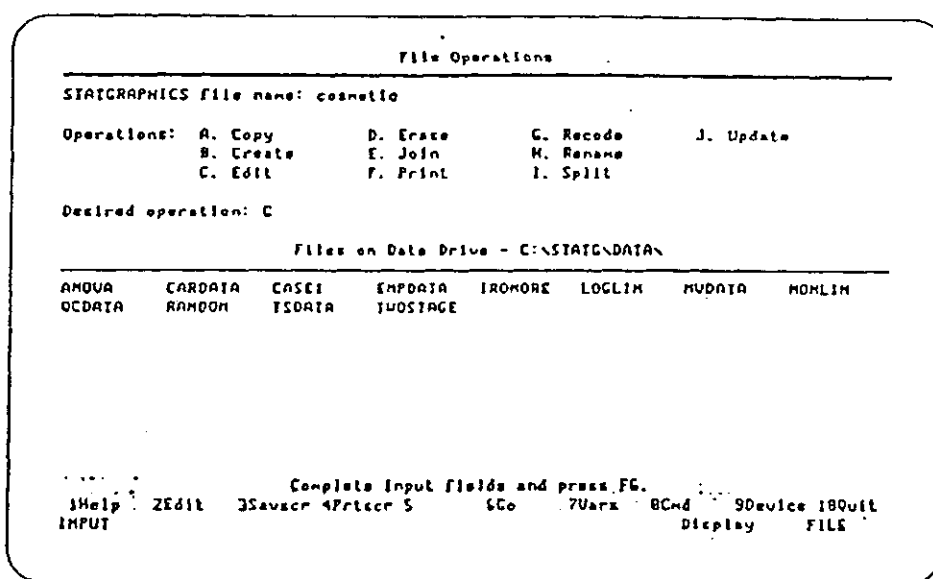


Figura 3-4. Pantalla de entrada de datos para Gestión de ficheros una vez completada

- Pulsa **F6 (Go)**.

Al pulsar **F6 (Go)** para procesar lo que se haya puesto en la pantalla de entrada de datos para Operaciones con ficheros, el sistema crea el fichero y entra en el editor de datos (véase la Figura 3-5). Observa que el sistema también muestra una ventana en blanco para Añadir otra columna. Una ventana es un recuadro que aparece en la pantalla y te pide que des más información. Al igual que en las pantallas de entrada de datos, las zonas seleccionadas en las ventanas se llaman campos.

Observa también el nombre del fichero, COSMETIC, en la parte superior central de la pantalla. Los nombres de fichero en STATGRAPHICS siempre aparecen en

mayúsculas; si se escribe en letras minúsculas, el sistema las convierte a mayúsculas : cosmetic a COSMETIC.

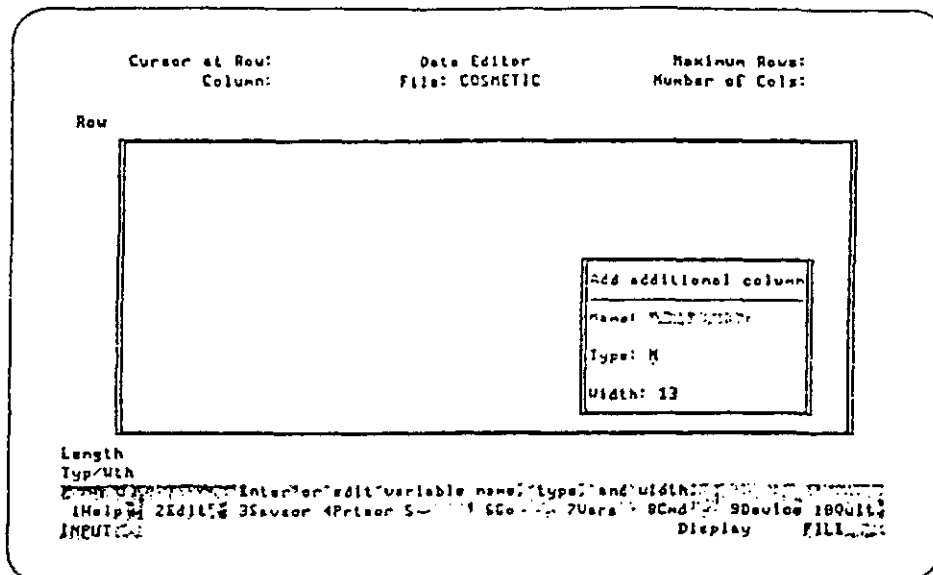


Figura 3-5. El Editor de datos de STATGRAPHICS con la ventana para Añadir otra columna

Ahora hay que crear una variable llamada **sales** (ventas) para poner en ella las ventas (en millones de dólares) de productos cosméticos. La ventana para Añadir otra columna se emplea para indicar el nombre de la variable- **sales**- y la amplitud de la nueva variable (véase la Figura 3-6). La amplitud de una variable es el máximo número de caracteres o dígitos que se quiera asignar a cada valor de datos.

Nota : A diferencia de los nombre de fichero de STATGRAPHICS, que siempre usan mayúsculas, los nombres de las variables pueden ponerse en mayúsculas, en minúsculas o combinando las dos. **SALES**, **Sales**, y **sales** son tres variables diferentes aunque contengan las mismas letras. Al escribir los nombres de las variables en esta guía, cerciórese de escribirlos **exactamente** en la forma indicada.

- Escribe las letras **sales** en el campo Name (Nombre).
- Pusa dos veces la tecla Tab para ir a campo Width (amplitud).
- Escribe el número 10 encima del número 13 en el campo Width

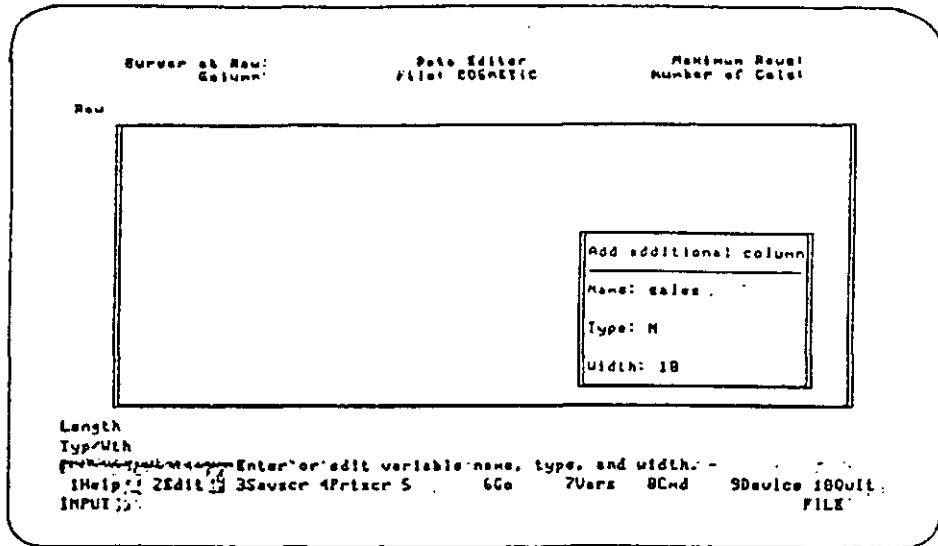


Figura 3-6. Ventana para Añadir otra columna una vez completada

- Pulsa **F6 (Go)**.

Al pulsar **F6 (Go)** para procesar lo entrado en la ventana para Añadir otra columna, el sistema crea la variable sales en el editor de datos y presenta una columna en blanco para usarla en la entrada de datos (véase la Figura 3-7).

Observa que el sistema también muestra una ventana nueva, en blanco, para Añadir otra columna.

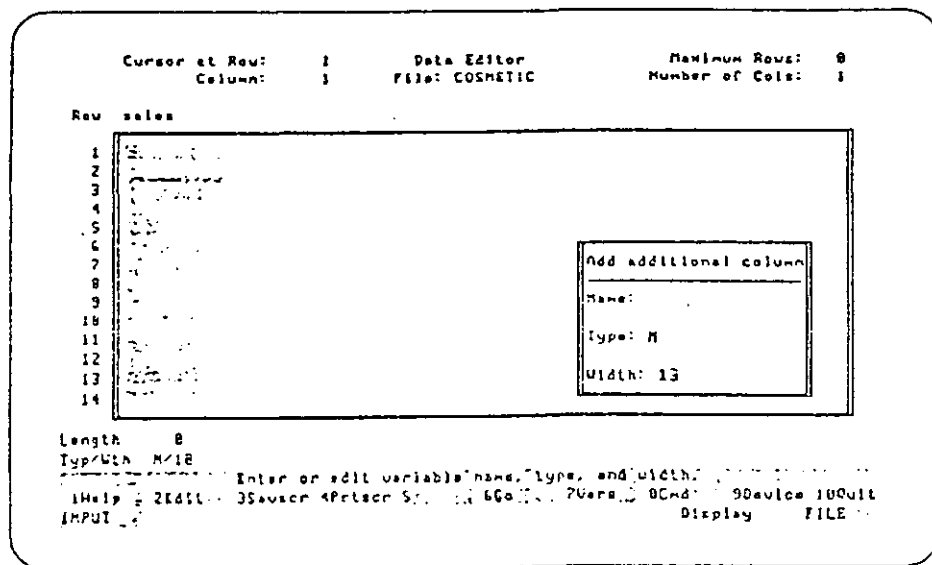


Figura 3-7. La columna de la variable sales en el editor de datos con la ventana para Añadir otra columna

El próximo paso es poner los valores de los datos en la columna de la variable sales. Pero antes hay que pulsar la tecla Esc para quitar de la pantalla la ventana para Añadir otra columna (véase la Figura 3-8).

- Pulsa la tecla Esc.

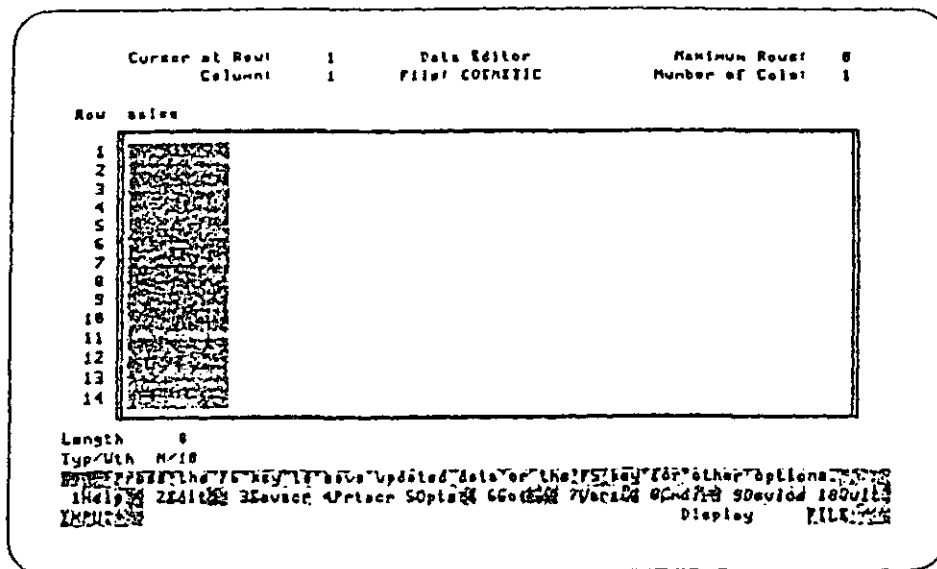


Figura 3-8. La columna de la variable sales en el editor de datos

Ahora estás preparado para entrar los valores de los datos (en millones de dólares) para la variable **sales** (véase la Figura 3-9). Como quieres que cada valor ocupe una celda (fila) de la columna de la variable, pulsa la tecla Enter después de escribir cada valor. Usa la tecla Backspace para corregir los errores.

- De ser necesario, usa la tecla Home para seleccionar la opción para Save Without Exit (Almacenar sin Salir).
- Escribe los valores indicados más abajo en la columna de la variable **sales**, pulsando la tecla Enter después de cada valor.

```
407 .293 .276 .274 .255 .253 .248 .245
240 .239 .223 .211 .200 .289
```

- Pulsa dos veces la tecla Home para volver a la primera celda.

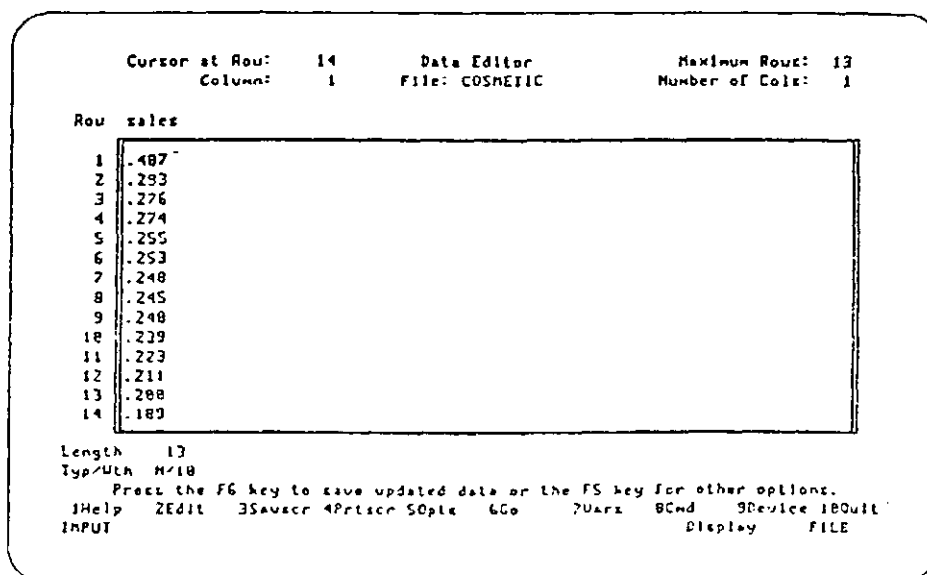


Figura 3-9. La variable sales en el editor de datos

- Pulsa F6 (Go).

Una vez se pulsa F6 (Go) para procesar las entradas en el editor de datos, el sistema presenta el menú-ventana para Almacenar. Como en el caso de la ventana, el menú-ventana ofrece dos o más opciones para que tú elijas una.

La opción Save Without Exit (Almacenar sin salir) del menú-ventana para Almacenar permite almacenar los datos y continuar dentro del editor de datos para poder añadir otras variables al fichero COSMETIC (véase la Figura 3-10).

- De ser necesario, usa la tecla Home para seleccionar la opción para Save Without Exit (Almacenar, sin Salir).

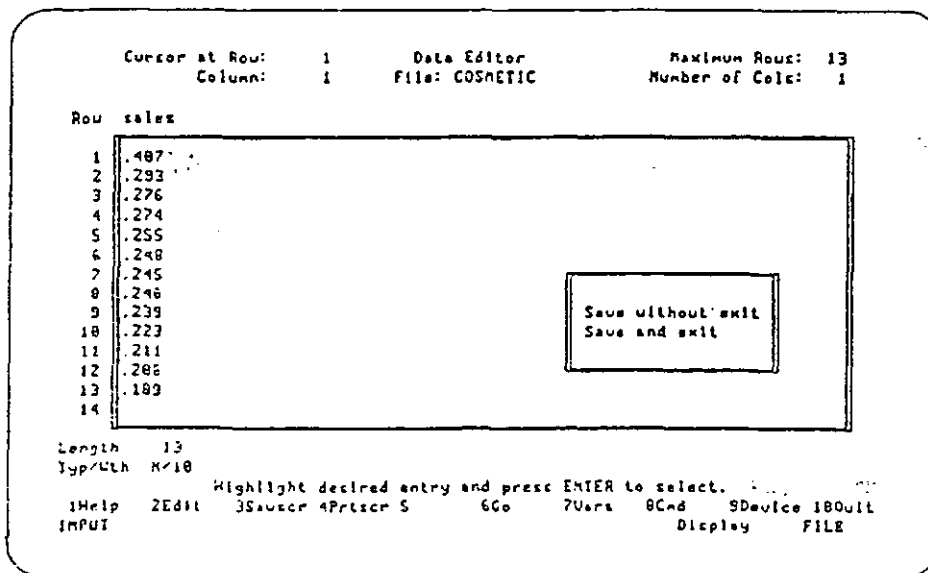


Figura 3-10. Menú-ventana para Almacenar, con selección de la opción Save Without Exit (Almacenar sin Salir)

- Pulsa la tecla Enter.



Una vez que se pulsa la tecla Enter para procesar la selección hecha en el menú-ventana para Almacenar, el sistema vuelve al editor de datos (véase la Fig. 3-11).

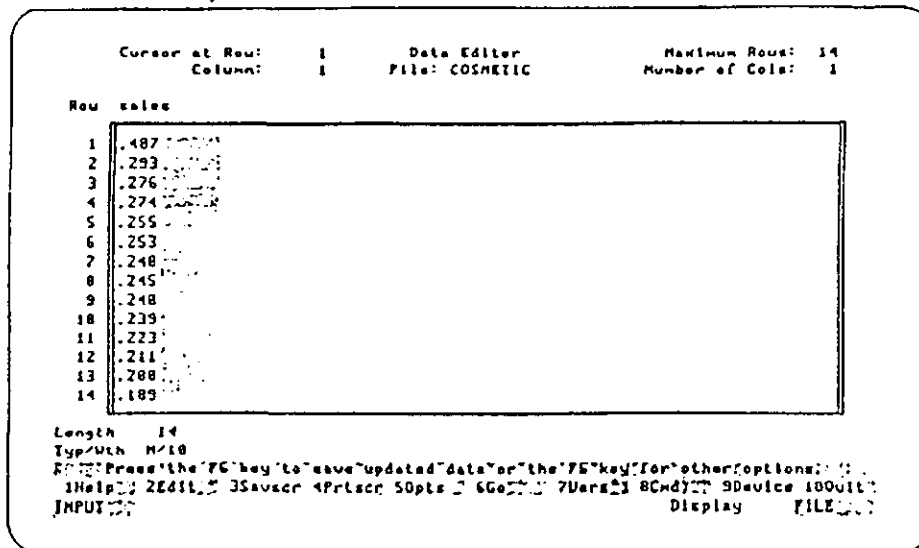


Figura 3-11. Editor de datos de STATGRAPHICS

Ahora hay que crear otra variable. Mira la línea de estado en la Figura 3-11. La tecla F5 aparece definida como Opts (Opciones). Pulsa **F5 (Opts)** para llegar al menú-ventana de Opciones del editor. En este menú-ventana selecciona la opción Add Additional Column (Añadir otra columna) para poder crear otra variable (véase la Figura 3-12).

- Pulsa **F5 (Opts)**.
- Usa la Flecha hacia abajo para seleccionar la opción para Añadir otra columna.

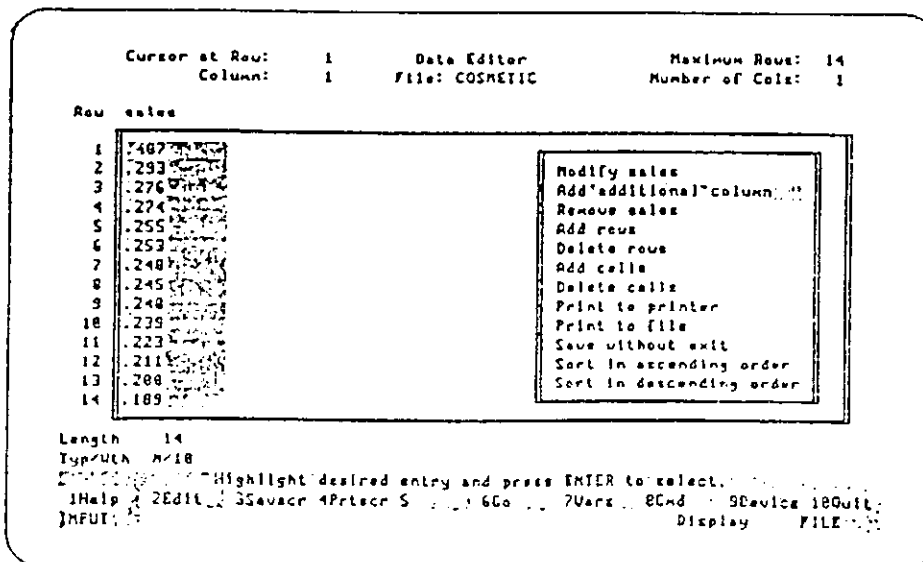


Figura 3-12. Menú-ventana de Opciones del editor con selección de la opción para Añadir otra columna

- Pulsa la tecla Enter.

Al apretar la tecla Enter para procesar la selección hecha en el menú-ventana de Opciones del editor, el sistema presenta una ventana en blanco para Añadir otra columna (véase la Figura 3-13).

Decimos que está "en blanco", aunque ya hay dos campos que tienen valores. ¿Por qué? Porque los valores que aparecen son opciones por defecto. Hay que indicar el nombre de la columna a añadir y, si se quiere, los valores de las opciones por defecto.

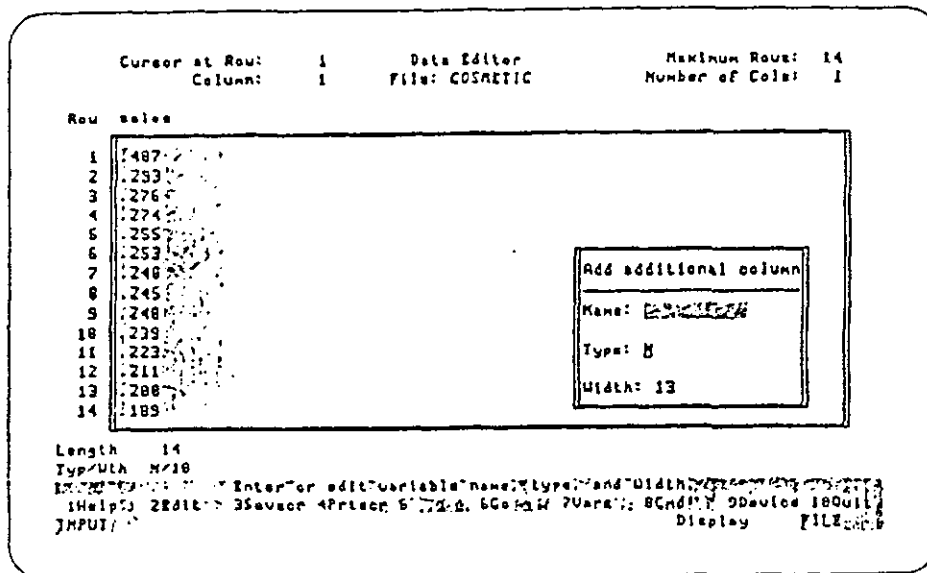


Figura 3-13. Ventana en blanco para añadir otra columna

Ahora estás preparado para crear una variable numérica (una variable que contenga números solamente) llamada **units**. La variable **units** tendrá las ventas (en miles de unidades) de productos cosméticos. Usa la ventana para Añadir otra columna al objeto de indicar el nombre de la nueva variable (véase la Fig. 3-14).

- Escribe la palabra **units** en el campo Name (Nombre).

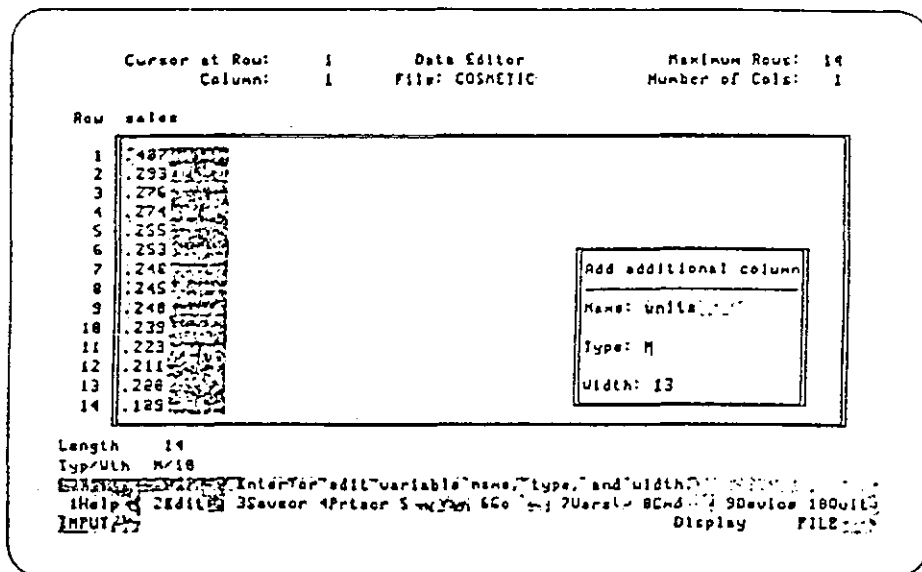


Figura 3-14. Ventana para Añadir otra columna una vez completada

- Pulsa F6 (Go).

Una vez que se pulsa F6 (Go) para procesar lo entrado en la ventana para Añadir otra columna, el sistema presenta una columna en blanco en el editor de datos, para contener los valores de los datos de **units**. Observa que el sistema también presenta una nueva ventana en blanco para Añadir otra columna (véase la Figura 3-15).

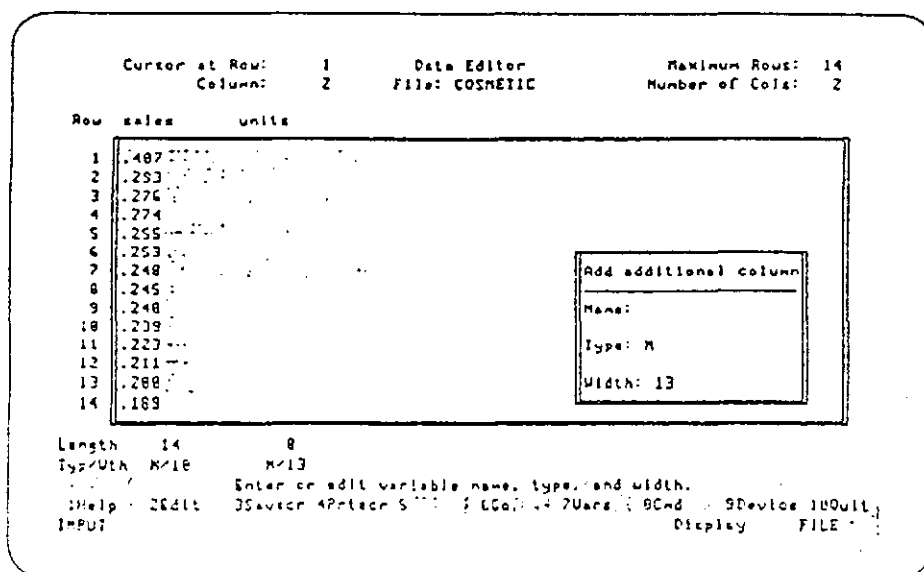


Figura 3-15. La columna de la variable units en el editor de datos con la ventana para Añadir otra columna

Antes de poner datos de la variable **units**, vamos a crear otra variable para la cantidad de inventario disponible (en miles de unidades) de productos cosméticos. Esta nueva variable, **onhand**, tendrá una amplitud de 10. Usa la ventana para Añadir otra columna al objeto de indicar el nombre de la nueva variable y su amplitud- el número máximo de dígitos que vayas a poner en cada valor de datos (véase la Figura 3-16).

- Escribe la palabra **onhand** en el campo Name (Nombre) de la ventana para Añadir otra columna.
- Pulsa la tecla **Tab** para pasar al campo Width (Amplitud).
- Escribe el número **10** encima del número 13 en el campo Width (amplitud).

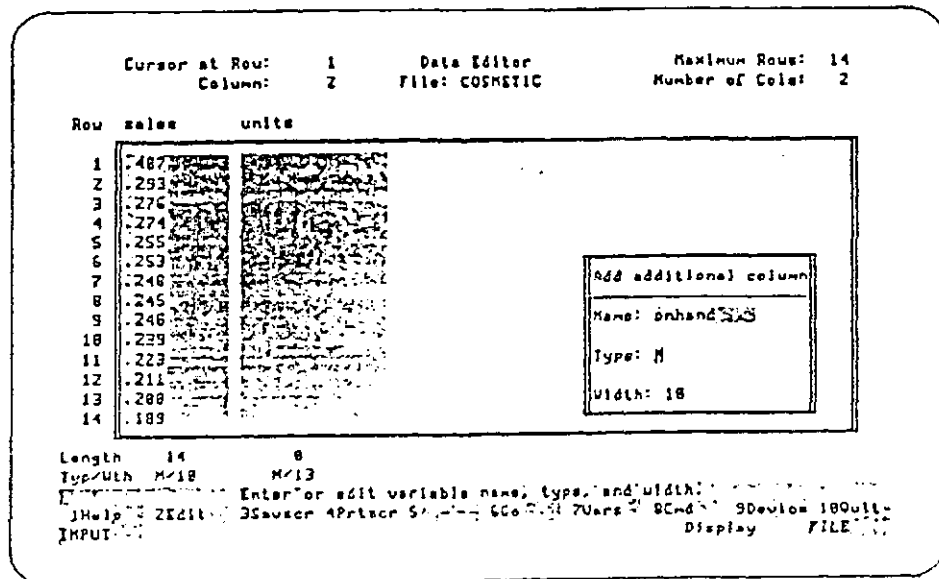


Figura 3-16. Ventana para Añadir otra columna una vez completada

- Pulsa **F6 (Go)**.

Una vez que se pulsa **F6 (Go)** para procesar lo entrado en la ventana para Añadir otra columna, el sistema crea la nueva columna, **onhand**, a la derecha de la columna de la variable **units** (véase la Figura 3-17). Observa que el sistema vuelve a presentar otra ventana en blanco para Añadir otra columna.

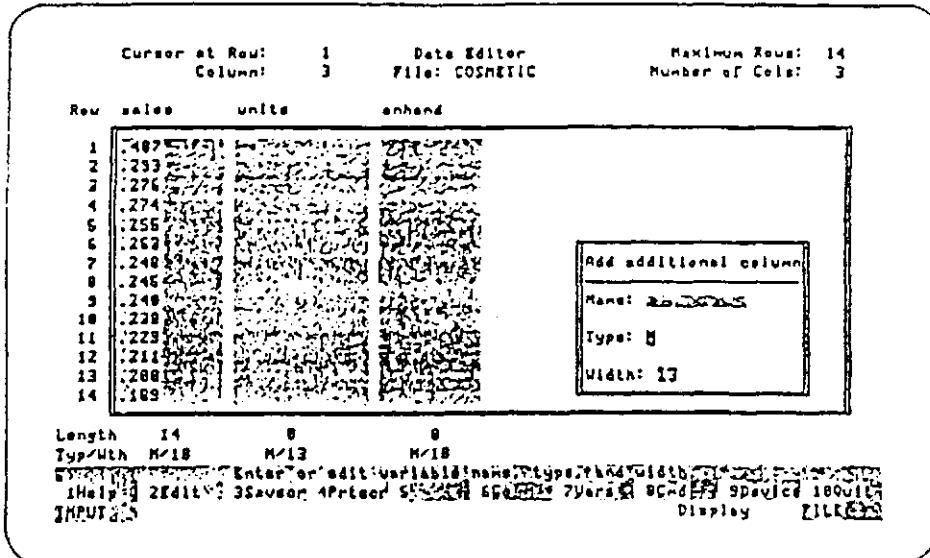


Figura 3-17. La columna de la variable onhand en el editor de datos con la ventana para Añadir otra columna

Ahora hay que poner los valores de los datos de las dos variables **units** y **onhand**. Pero antes tienes que usar la tecla **Esc** para quitar de la pantalla la ventana para Añadir otra columna (véase la Figura 3-18).

- Pulsa la tecla **Esc**.

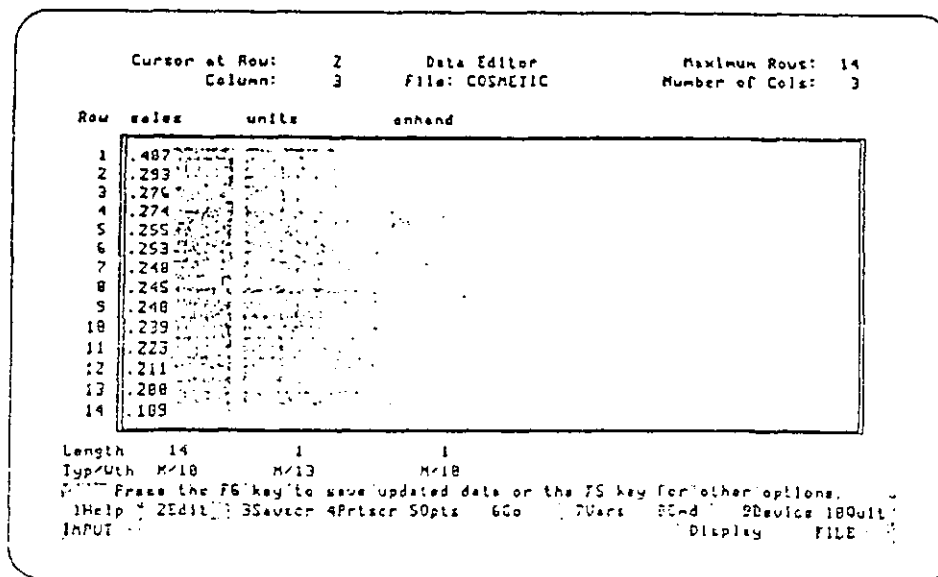


Figura 3-18. Editor de datos de STATGRAPHICS

Ahora puedes meter datos en las variables **units** y **onhand** (véase la Fig. 3-19 ). Usa la tecla Backspace para corregir errores. Pulsa la tecla Enter después de escribir cada número.

- Usa las teclas Home y Tab para poner el cursor en la primera celda (fila) de la columna de la variable **units**.

Escribe los valores siguientes en la columna de la variable **units**, y pulsa la tecla Enter después de cada valor.

19 12 8 13 22 11 41 35 38 22 26 30 50 47

Usa la Flecha hacia arriba y la tecla Tab para poner el cursor a la izquierda de la primera celda de la variable **onhand**.

Escribe los valores siguientes en la columna de la variable **onhand**, y pulsa la tecla Enter después de cada valor.

9 61 67 56 56 70 63 29 41 59 63 33 4 32

Cursor at Row: 14  
Column: 3  
Data Editor  
File: COSMETIC  
Maximum Row: 14  
Number of Col: 3

Row	sales	units	onhand
1	487	19	9
2	293	12	61
3	276	8	67
4	274	13	56
5	265	22	56
6	262	11	70
7	248	35	63
8	245	38	29
9	245	22	41
10	239	26	59
11	223	30	63
12	211	50	33
13	208	47	4
14	192	22	32

Length 14 14 17  
Typ/Uch N/18 N/13 N/18  
Display FILE

Figura 3-19. Las variables **units** y **onhand** en el editor de datos

- Pulsa F6 (Go)

Una vez que se pulsa **F6 (Go)** para procesar lo entrado en el editor de datos, el sistema presenta el menú-ventana **Save (Almacenar)**. Ahora vas a almacenar los datos y a salir del editor de datos. Selecciona la opción **Save and Exit (Almacenar y Salir)** en el menú-ventana para Almacenar (véase la Figura 3-20).

- Usa la Flecha hacia abajo para seleccionar la opción **Save and Exit (Almacenar y Salir)**

Row	sales	units	onhand
1	.487	18	9
2	.293	12	61
3	.276	8	67
4	.274	13	66
5	.255	22	55
6	.253	11	70
7	.249	41	63
8	.245	35	29
9	.248	38	41
10	.239	22	59
11	.223	26	63
12	.211	38	33
13	.268	58	4
14	.189	47	32

Length	14	14	14
Input	M/18	M/13	M/10
Highlight desired entry and press ENTER to select.			
1Help	2Edit	3Save	4Print
5	6Go	7Quit	8End
9Device	10Quit	Display	FILE

Figura 3-20. Menú-ventana **Save (Almacenar)** con selección de la opción **Save and Exit (Almacenar y Salir)**

- Pulsa la tecla **Enter**

Una vez que se pulsa la tecla **Enter** para procesar la selección en el menú-ventana para Almacenar, el sistema almacena los datos y sale del editor de datos. Después, el sistema presenta la pantalla **File Contents (Contenido del fichero)** del fichero **COSMETICS** (véase la Figura 3-21). En la pantalla se ven todas las variables del fichero **COSMETICS**.

La longitud de una variable es el número de filas en la columna de la variable- el número de valores que contiene la variable. Cada una de estas variables tiene una longitud de 14. (Si tienes más filas después de la 14, mira las instrucciones en la Lección 3-3 para borrarlas)

Ahora queremos ver el contenido de la variable **sales**. Mira la línea del estado. La opción D (Display-Visualización) te permite ver el contenido de una variable. Pero antes tienes que seleccionar la variable en la pantalla de Contenido de ficheros.

- Pulsa la tecla Home, si hace falta, para seleccionar la variable **sales**.

The following variables are currently in the file COSMETIC:

Variable	Width	Type	Rank	Length	Date	Time	Comment
sales	18	M	1	14	11/12/98	11:13	
units	13	M	1	14	11/12/98	11:13	
onhand	18	M	1	14	11/12/98	11:13	

Press ENTER key for each variable to be edited, or press A=All D=Display. ...  
 1)Help 2)Edit 3)Savecr 4)Prtnr 5) 6)Go 7)Arc 8)Cmd 9)Device 10)Quit  
 INPUT Display FILE

Figura 3-21. Pantalla de Contenido de fichero con el fichero COSMETIC con la variable **sales** seleccionada

- Escribe la letra D (opción Display-visualizar).

Una vez escrita la letra D (opción de visualización) en la pantalla de Contenido de ficheros, el sistema presenta el contenido de la variable **sales**. La pantalla de Contenido que se muestra en la Figura 3-22 presenta el nombre de la variable, su longitud y los valores de los datos. Desde esta pantalla no se puede editar los valores almacenados en la variable.

En la Lección 3-2 aprenderás cómo crear una variable de caracteres - una variable en la que puede haber letras y números. La lección 3-2 empieza en la pantalla de Contenido de ficheros, así que usa la tecla Esc para volver a la pantalla de Contenido de ficheros.

- Pulsa la tecla Esc para volver a la pantalla de Contenido de fichero.



```
Variable: COSMETIC.sales (length = 14)
-----
< 1> 8.487
< 2> 8.293
< 3> 8.276
< 4> 8.274
< 5> 8.255
< 6> 8.253
< 7> 8.248
< 8> 8.245
< 9> 8.24
<10> 8.239
<11> 8.223
<12> 8.211
<13> 8.2
<14> 8.189
-----

Press Esc, Cursor keys or Page Number:          Page 101 of 101
[Help] [2Edit] [3Save] [4Print] [5Print] [6Co] [7Vars] [8Cmd] [9Device] [0Quit]
INPUT$          Display          FILE$
```

Figura 3-22. Pantalla de Contenido de la variables sales

### CREACION DE UNA VARIABLE DE CARACTERES.

En esta lección aprenderás cómo crear una variable de caracteres llamada **products** usando el editor de datos. Almacenarás la variable en el fichero **COSMETIC** que consiste en la Lección 3-1.

La Lección 3-2 empieza en la pantalla de Contenido de ficheros.

La pantalla de Contenido de ficheros indica todas las variables que hay en el fichero **COSMETIC** (véase la Figura 3-23). Queremos ahora ver todas las variables en el editor de datos. Mira la Línea de estado. La opción **A** (All -Todas) te permite ver el contenido de todas las variables.

- Escribe la letra **A** (opción All-Todas).

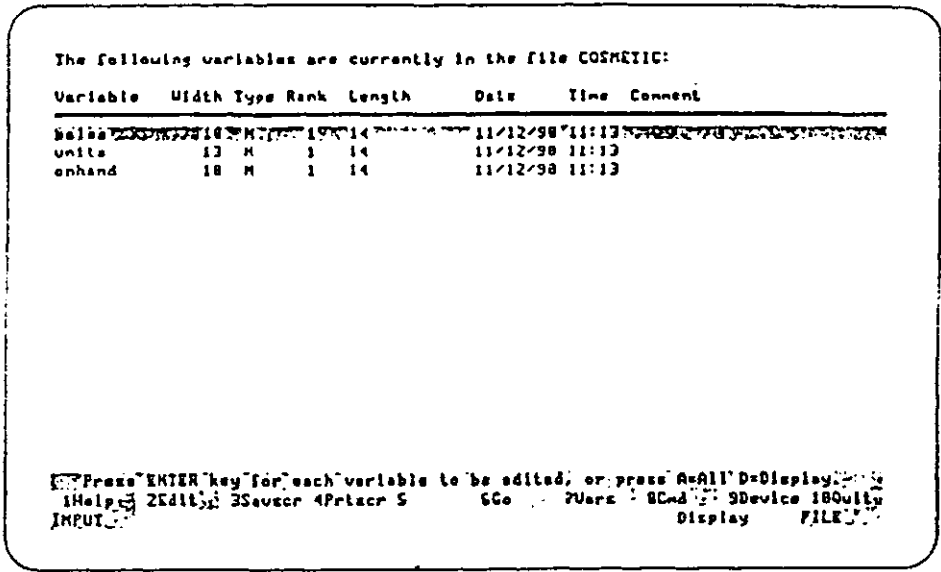


Figura 3-23. Pantalla de Contenido del fichero con la variable sales

Una vez que escriba la letra A (para la opción All-Todas) en la pantalla de Contenido de ficheros, el sistema presenta las variables sales, units, y onhand en el editor de datos (véase la Figura 3-24).

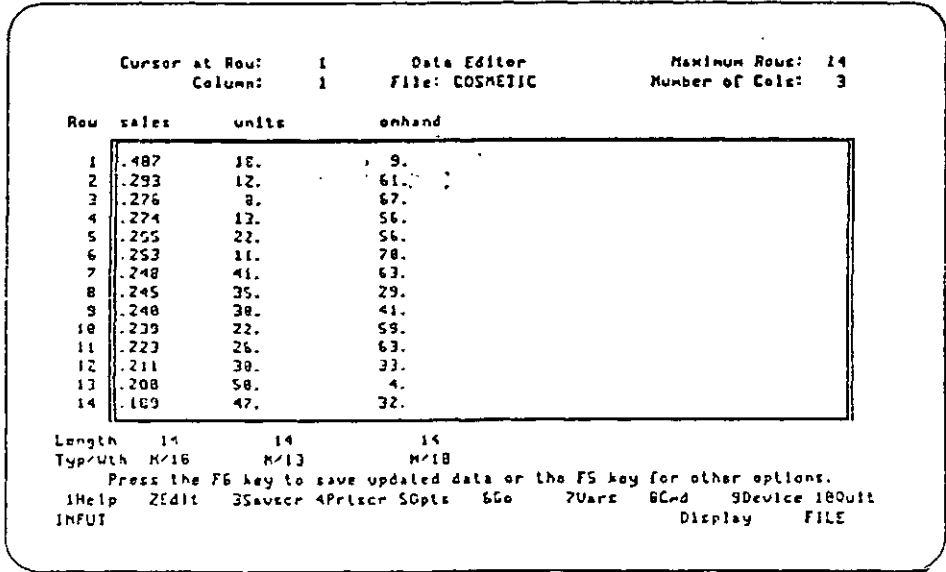


Figura 3-24. Las variables sales, units, y onhand en el editor de datos

Ahora tienes que crear otra variable- **products**- a la derecha de la variable **onhand**. La variable **products** tendrá los nombres de los 14 productos cosméticos. Observa la línea de estado. La tecla **F5 (Opts)** (Opciones) te permite ver otras opciones. Pero antes de que uses **F5 (Opts)** tienes que poner el cursor en la columna de la variable **onhand**.

- Usa la tecla Tab para poner el cursor en la columna de la variable **onhand**.
- Pulsa **F5 (Opts)**.

Una vez que se pulsa **F5 (Opts)**, el sistema presenta el menú-ventana de Opciones del editor. Para crear una variable nueva, tienes que seleccionar la opción para Añadir otra columna, en el menú-ventana de Opciones del editor (véase la Figura 3-25).

- Usa la Flecha hacia abajo para seleccionar la opción Add Additional Column (Añadir otra columna).

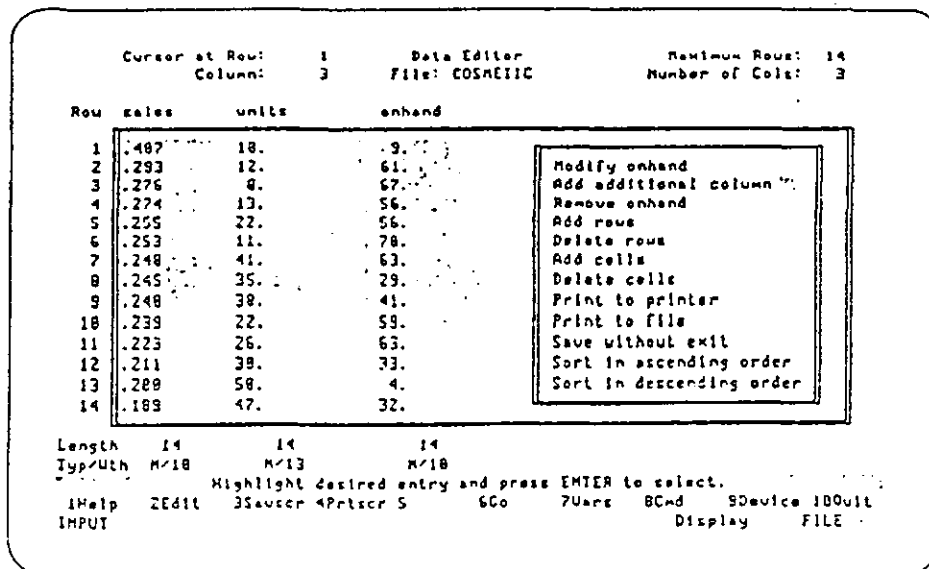


Figura 3-25. Menú-ventana de Opciones del editor con selección de la opción Add Additional Column (Añadir otra columna)

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar la selección hecha en el menú-ventana de Opciones del editor, el sistema presenta la ventana para Añadir otra columna (véase la Figura 3-26).

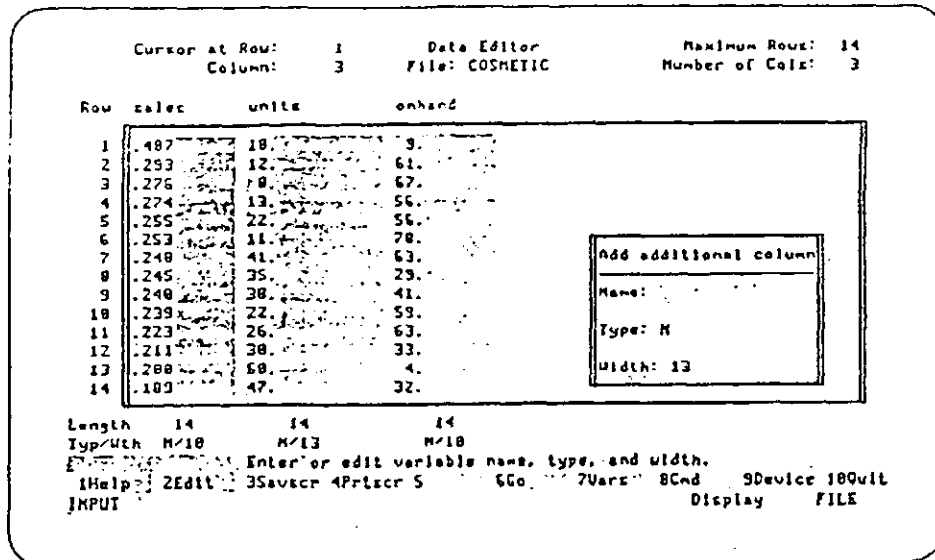


Figura 3-26. Editor de datos con la ventana Add Additional Column (Añadir otra columna)

Para crear la variable **products**, se indica el nombre de la nueva variable en la ventana para Añadir otra columna (véase la Figura 3-27). También se usa la misma ventana para indicar que **products** es una variable de caracteres con una amplitud de 30.

- Escribe la palabra **products** en el campo Name (Nombre) de la ventana Add Additional Column (Añadir otra columna).
- Usa la tecla Tab para poner el cursor en el campo Type (Tipo).
- Escribe la letra C (Caracteres) en el campo Type.
- Usa la tecla Tab para poner el cursor en el campo Width (Amplitud).
- Escribe el número 30 encima del número 13 en el campo width (Amplitud).

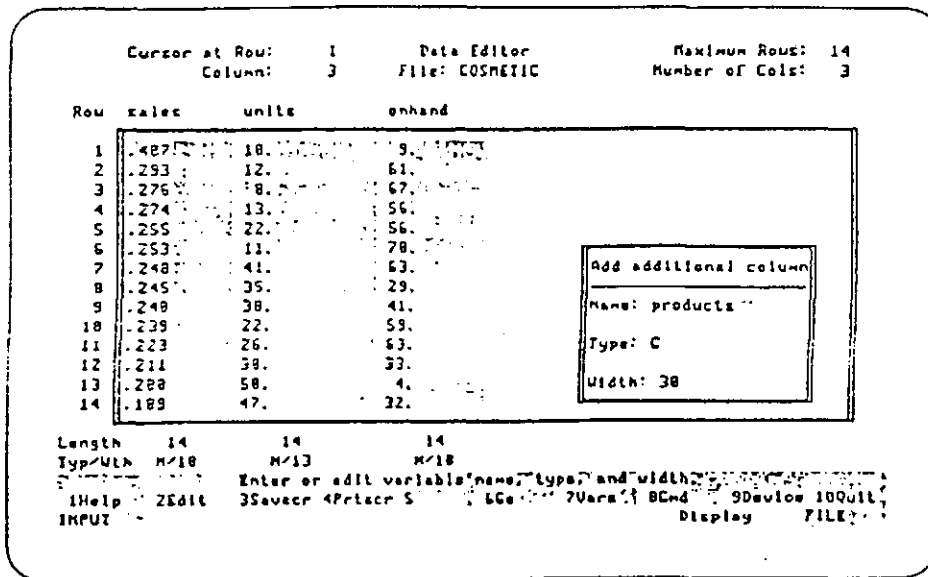


Figura 3-27. Ventana para Añadir otra columna una vez completada

- Pulsa **F6 (GO)**.

Una vez que se pulsa **F6 (Go)** para procesar lo entrado en la ventana para Añadir otra columna, el sistema crea una columna en blanco para la nueva variable (véase la Figura 3-28). Observa que el sistema también presenta una ventana en blanco para Añadir otra columna.

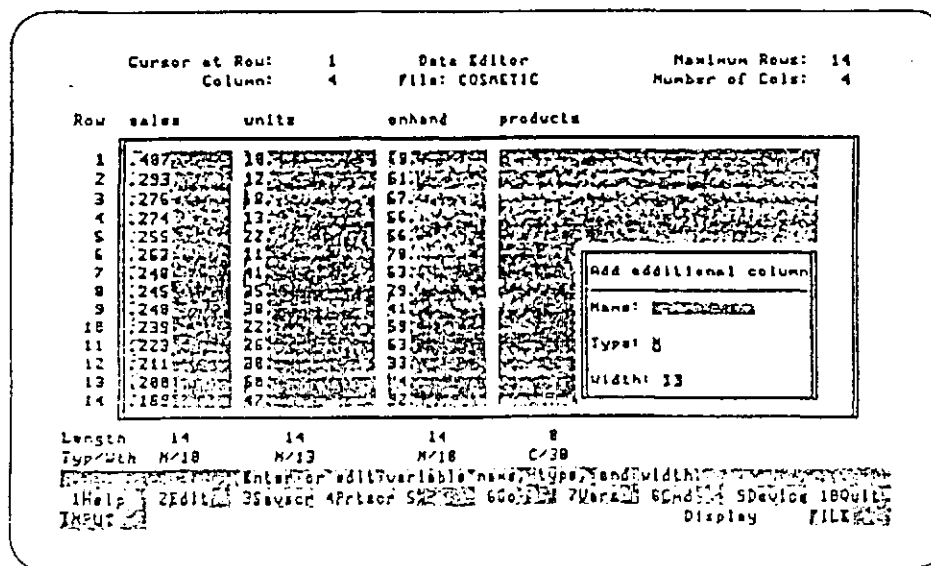


Figura 3-28. La columna de la variable products en el editor de datos y la ventana Add Additional Column (Añadir otra columna)

Ahora hay que poner valores de datos en la nueva variable **products**. Pero antes hay que usar la tecla Esc para hacer desaparecer de la pantalla la ventana para Añadir otra columna (véase la Figura 3-29).

- Pulsa la tecla Esc.

Cursor at Row: 1      Data Editor      Maximum Row: 14  
 Column: 4      File: COSMETIC      Number of Col: 4

Row	sales	units	onhand	products
1	.487	18	59	
2	.293	12	61	
3	.276	19	67	
4	.274	13	66	
5	.255	22	56	
6	.253	11	78	
7	.248	41	63	
8	.245	35	23	
9	.240	39	41	
10	.239	22	59	
11	.223	26	63	
12	.211	39	33	
13	.208	58	14	
14	.189	47	32	

Length 14      14      14      8  
 Typ/Wth N/18    N/13    N/18    C/38  
 Press the F6 key to save updated data or the F5 key for other options:  
 1Help 2Edit 3Save 4Print 5Opts 6Go 7Vars 8Cmd 9Device 10Quit  
 INPUT      Display    FILE

Figura 3-29. La columna de la variable products en el editor de datos

Ahora todo está preparado para entrar valores de datos para la nueva variable **products**. Puedes emplear las teclas de las flechas (arriba, abajo, derecha, e izquierda) y las teclas Tab, Home, y End además de la tecla Enter para ir de una parte a otra dentro del editor de datos. Usa la tecla Backspace (Retroceso) para corregir errores. Acuérdate de pulsar la tecla Enter después de escribir cada uno de los nombres (véase la Figura 3-30).

- Escribe los nombres que se indican a continuación en la columna de la variable **products**.

Madness Perfume  
 Sweetness Perfume  
 Roses Perfume  
 Sea Foam Cologne  
 Madly Musk Cologne  
 Simmering Spice Cologne  
 Dusty Dawn Lipstick  
 Sultry Sunset Lipstick  
 Red Red Rose Lipstick  
 Hint of Rose Powder  
 Basically Beige Powder  
 Plainly Peach Powder  
 Ocean Blue Shadow  
 Barely Brown Shadow

Cursor at Row: 14      Data Editor      Maximum Row: 14  
 Column: 4      File: COSMETIC      Number of Col: 4

Row	sales	units	onhand	products
1	.487	18.	79.	Madness Perfume
2	.293	12.	61.	Sweetness Perfume
3	.276	9.	67.	Roses Perfume
4	.274	13.	56.	Sea Foam Cologne
5	.255	22.	56.	Madly Musk Cologne
6	.253	11.	78.	Simmering Spice Cologne
7	.248	41.	63.	Dusty Dawn Lipstick
8	.245	35.	29.	Sultry Sunset Lipstick
9	.248	38.	41.	Red Red Rose Lipstick
10	.239	22.	59.	Hint of Rose Powder
11	.223	26.	63.	Basically Beige Powder
12	.211	38.	33.	Plainly Peach Powder
13	.288	58.	74.	Ocean Blue Shadow
14	.189	47.	32.	Barely Brown Shadow

Length 14      14      14      13  
 TypPrch M/18      M/13      M/18      C/38  
 Press the F6 key to save updated data or the F5 key for other options.  
 Help 2Edit 3Savscr 4Prtscr 5Opta 6Co 7Uars 8C-d 9Device 18Quit  
 INPUT      Display      FILE

Figura 3-30. La variable products en el editor de datos

- Pulsa F6 (Go).

Una vez que se pulsa F6 (Go) para procesar las entradas en el editor de datos, el sistema presenta el menú-ventana para Almacenar. Hay que almacenar los datos y salir del editor de datos. Selecciona la opción Save and Exit (Almacenar y salir) en el menú-ventana para Almacenar (véase la Figura 3-31).

- Usa la Flecha hacia abajo para seleccionar la opción Save and Exit (Almacenar y Salir).

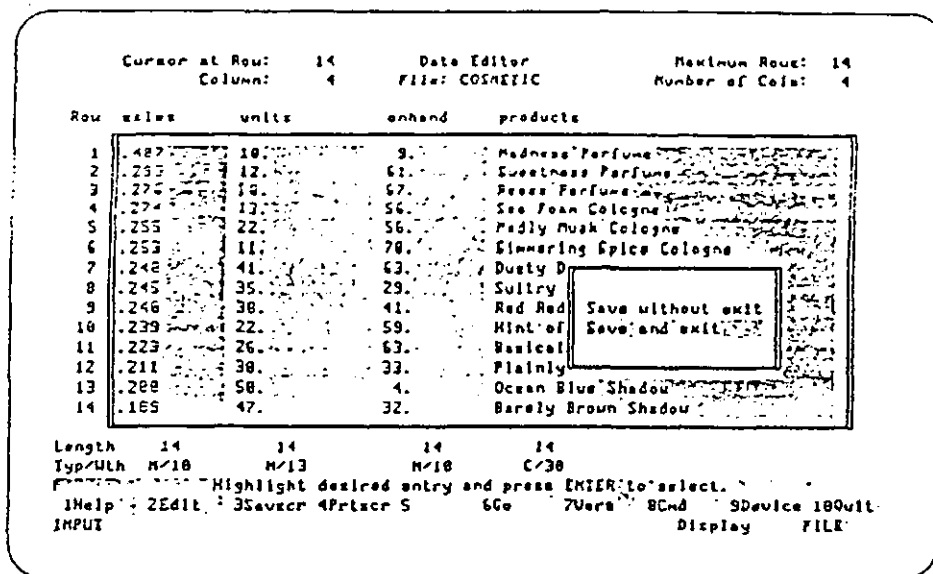


Figura 3-31. Menú-ventana Save (Almacenar) con selección de la opción Save and Exit (Almacenar y Salir)

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo seleccionado en el menú-ventana para Almacenar, el sistema almacena los datos y sale del editor de datos, y luego presenta la pantalla de Contenido de ficheros. Observa que la pantalla de Contenido de ficheros ahora presenta información sobre la nueva variable products.

A continuación vamos a examinar el contenido de la variable products. Se usa la opción D (Display- Visualización) de la pantalla de Contenido de ficheros para ver la variable products en el editor de datos. Antes hay que seleccionar la variable products en la pantalla de Contenido de Ficheros (véase la Figura 3-32).

- Usa la tecla End para seleccionar la variable products.



```

The following variables are currently in the file COSMETIC:

Variable  Width Type Rank Length      Date      Time Comment
-----
sales      18  M   1   14      11/12/98 12:56
units      13  M   1   14      11/12/98 12:56
onhand     18  M   1   14      11/12/98 12:56
products   38  C   2  14 38      11/12/98 12:56

Press ENTER key for each variable to be edited, or press A=All D=Display.
1Help  2Edit  3Savecr 4Prtsr 5      6Co   7Users  8Cmd  9Device 10Out
INPUT                                     Display FILE

```

Figura 3-32. Pantalla de Contenido de fichero del fichero COSMETIC con selección de la variable products - Escribe la letra D (para visualizar)

Una vez que se pulsa D (para la opción de visualización), el sistema muestra la pantalla de Contenido con la variable **products** (véase la Figura 3-33). La pantalla de Contenido presenta el nombre de la variable, su longitud (número de valores y amplitud permitida para cada valor), y los valores de los datos. No se pueden editar los valores de los datos en la variable desde esta pantalla de Contenido.

```

Variable: COSMETIC.products (length = 14 38)
-----
< 1)Madness Perfume
< 2>Sweetness Perfume
< 3>Roses Perfume
< 4>Sea Foam Cologne
< 5>Madly Musk Cologne
< 6>Simmering Spice Cologne
< 7>Dusty Dawn Lipstick
< 8>Sultry Sunset Lipstick
< 9>Red Red Rose Lipstick
<10>Hint of Rose Powder
<11>Basically Beige Powder
<12>Plainly Peach Powder
<13>Ocean Blue Shadow
<14>Barely Brown Shadow
-----

Press 'Esc' Cursor keys or Page Number:
1Help  2Edit  3Savecr 4Prtsr 5Prpt 6Go 7Users 8Cmd 9Device 10Out
INPUT                                     Display FILE

```

Figura 3-33. Pantalla de Contenido de la variable products

En la Lección 3-3 aprenderás cómo editar las variables existentes. La Lección 3-3 empieza en la pantalla de Contenido de ficheros, así que usa la tecla Esc para volver a la pantalla de Contenido de ficheros (véase la Figura 3-34).

- Pulsa la tecla Esc para volver a la pantalla de Contenido de fichero.

The following variables are currently in the file COSMETIC:

Variable	Width	Type	Rank	Length	Date	Time	Comment
sales	18	N	1	13	11/ 8/98	17:07	
units	13	N	1	14	11/ 8/98	17:07	
onhand	18	N	1	14	11/ 8/98	17:07	
products	38	C	2	14 38	11/ 8/98	17:07	

Press ENTER key for each variable to be edited, or press A=All D=Display.  
 1Help 2Edit 3Save 4Print 5Go 6Vars 7Cmd 8Device 9Quit  
 INPUT Display FILE

Figura 3-34. Pantalla de Contenido del fichero COSMETIC con la variable products seleccionada

## EDICION DE VARIABLES EXISTENTES.

En esta lección aprenderás cómo editar las variables creadas en las Lecciones 3-1 y 3-2 empleando el editor de datos.

La lección 3-3 empieza en la pantalla de Contenido de ficheros.

La pantalla de Contenido de ficheros ofrece la lista de todas las variables del fichero COSMETICS (véase la Figura 3-35).

- Escribe la letra A (opción All-Todas).

The following variables are currently in the file COSMETIC:

Variable	Width	Type	Rank	Length	Date	Time	Comment
sales	18	N	1	13	11/	8/98	17:07
units	13	N	1	14	11/	8/98	17:07
onhand	18	N	1	14	11/	8/98	17:07
products	18	N	2	14	11/	8/98	17:07

Press ENTER key for each variable to be edited, or press A=All D=Display.  
1=Help 2=Edit 3=Sever 4=Protect 5 Go 7=User 8=Cmd 9=Device 10=Quit  
INPUT Display FILE

Figura 3-35. Pantalla de Contenido del fichero COSMETIC

Una vez que se escribe la letra A (All-Todas) en la pantalla de Contenido de ficheros, el sistema presenta las variables **sales**, **units**, **onhand**, y **products** en el editor de datos (véase la Figura 3-36). Observa que el sistema presenta el contenido de las variables en el orden en que figuraban en la pantalla de Contenido de ficheros.

Row	sales	units	onhand	products
1	.487	18.	9.	Madness Perfume
2	.293	12.	61.	Sweetness Perfume
3	.276	8.	67.	Roses Perfume
4	.274	13.	56.	Sea Foam Cologne
5	.255	22.	56.	Nadiy Musk Cologne
6	.253	11.	78.	Simmering Spice Cologne
7	.248	41.	63.	Dusky Dawn Lipstick
8	.245	35.	29.	Sultry Sunset Lipstick
9	.248	38.	41.	Red Red Rose Lipstick
10	.239	22.	59.	Hint of Rose Powder
11	.223	26.	63.	Basically Beige Powder
12	.211	38.	33.	Plainly Peach Powder
13	.208	58.	4.	Ocean Blue Shadow
14	.189	47.	32.	Rarely Brown Shadow

Length	14	14	14	14
Typ/Wth	N/18	N/13	N/18	C/38

Press the F6 key to save updated data or the F5 key for other options.

1H=lp 2Edit 3Savecr 4Prtscr 5Opts 6Go 7Ware 8Cnd 9Device 10Quit  
 INPUT Display FILE

Figura 3-36. Las variables sales, units, onhand, y products en el editor de datos

Ahora se quiere cambiar los valores de los datos en las celdas de la fila 14. Se puede editar las variables simplemente escribiendo de nuevo los valores de los datos (véase la Figura 3-37).

- Usa la tecla End para poner el cursor en la celda 14 de la variable sales. Escribe el número .154 encima del número .189.
- Usa la tecla Tab para ir a la celda 14 de la variable units. Escribe el número 32 encima del número 47.
- Usa la tecla Tab para ir a la celda 14 de la variable onhand. Pon un espacio y un número 9 encima del número 32.
- Usa la tecla Tab para ir a la celda 14 de la variable products. Escribe el nombre Silvery Shadow encima del nombre Rarely Brown Shadow. Usa la tecla Delete (Borrar) para borrar las letras sobrantes.

Row	sales	units	onhand	products
1	.487	18.	9.	Madness Perfume
2	.293	12.	61.	Sweetness Perfume
3	.276	18.	67.	Roses Perfume
4	.274	13.	56.	Sea Foam Cologne
5	.255	22.	56.	Madly Musk Cologne
6	.253	11.	78.	Simmering Spice Cologne
7	.248	41.	63.	Dusty Dawn Lipstick
8	.245	35.	29.	Sultry Sunset Lipstick
9	.248	36.	41.	Red Red Rose Lipstick
10	.239	22.	53.	Mint of Rose Powder
11	.223	26.	63.	Esotically Esige Powder
12	.211	38.	33.	Plainly Peach Powder
13	.208	58.	4.	Ocean Blue Shadow
14	.154	32.	9.	Silvery Shadow

Length 14 14 14 14  
 Typ/Uch M/18 M/13 M/18 C/38  
 \* Press the F5 key to save updated data or the F5 key for other options.  
 !Help 2Edit 3Savecr 4Prtscr 5Opts 6Go 7Ware 8Cnd 9Device 10Outt  
 INPUT... Display FILE

Figura 3-37. Variables editadas en el editor de datos

A continuación se quiere borrar una fila entera de valores de datos. Primero hay que seleccionar la fila a borrar. Luego se pulsa **F5 (Opts)** para abrir el menú-ventana de Opciones del editor. Para terminar, se selecciona la opción Delete Rows (Borrar filas) en el menú-ventana de Opciones del editor (véase la Figura 3-38 ).

- Usa la tecla Home para poner el cursor al principio de la columna de la variable sales.
- Usa la Flecha hacia abajo para poner el cursor en la celda 5 (valor de datos .255).
- Pulsa **F5 (Opts)**.
- Usa la Flecha hacia abajo para seleccionar la opción Delete Rows (Borrar filas).

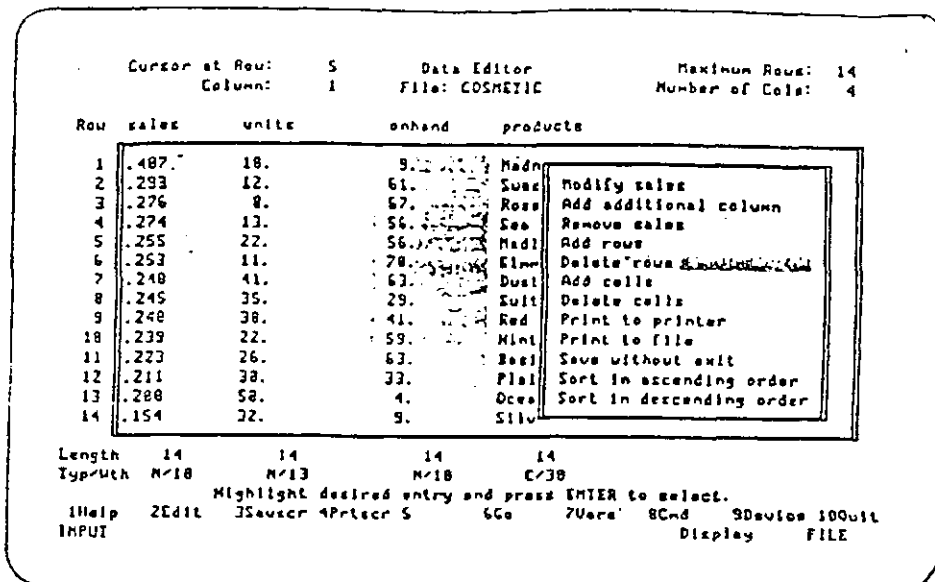


Figura 3-38. Menú-ventana de Opciones del editor con selección de la opción Delete Rows (Borrar filas)

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo seleccionado en el menú-ventana de Opciones del editor, el sistema pide que se indique el número de filas a borrar (véase la Figura 3-39).

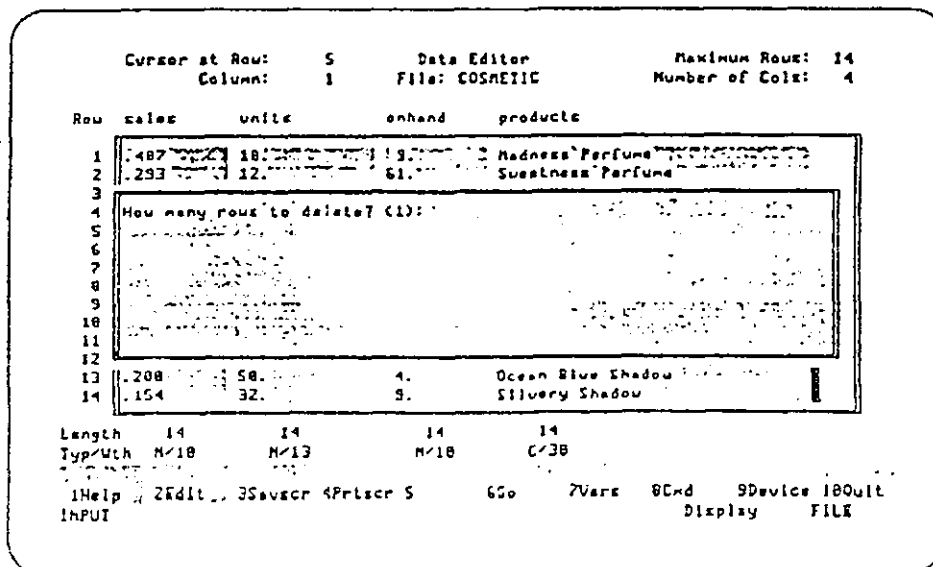


Figura 3-39. Indicador del número de filas a borrar

Hay que borrar una fila de las variables **sales**, **units**, **onhand**, y **products**.

- Pulsa la tecla Enter para aceptar el valor de 1 fila a borrar como opción por defecto.

Una vez que se pulsa la tecla Enter para aceptar la opción por defecto con valor de 1, el sistema borra la fila 5 y hace avanzar un puesto las filas siguientes (véase la Figura 3-40). Observa que cuando se emplea la opción para Borrar filas, STATGRAPHICS empieza a borrarlas a partir de la fila en la que se encuentre el cursor.

```
Cursor at Row: 5      Data Editor      Maximum Rows: 13
                Column: 1      File: COSMETIC      Number of Cols: 4
```

Row	sales	units	onhand	products
1	.487	18.	9.	Hedness Perfume
2	.293	12.	61.	Sweetness Perfume
3	.276	18.	67.	Roses Perfume
4	.274	13.	56.	Sea Foam Cologne
5	.253	11.	78.	Simmering Spice Cologne
6	.248	41.	63.	Dusty Dawn Lipstick
7	.245	35.	29.	Sultry Sunset Lipstick
8	.248	38.	41.	Red Red Rose Lipstick
9	.239	22.	59.	Mint of Rose Powder
10	.223	26.	63.	Basically Beige Powder
11	.211	38.	33.	Plainly Peach Powder
12	.288	58.	4.	Ocean Blue Shadow
13	.154	32.	9.	Silvery Shadow
14				

```
Length 13      13      13      13
Typ/VLH N/10   N/13   N/18   C/30
Press the F6 key to save updated data or the F5 key for other options.
INHelp 2Edit 3Savecr 4Prtscr 5Opts 6Go 7Vars 8Cmd 9Device 18Quit
INPUT                                     Display FILE
```

Figura 3-40. Variables con una fila borrada en el editor de datos

A continuación hay que añadir una fila entera de valores de datos. Primero hay que indicar dónde se quiere añadir la fila. (Cuando se usa la opción para Añadir filas, STATGRAPHICS empieza a añadirlas en la fila donde se encuentra el cursor). Luego se pulsa F5 (Opts) para abrir el menú ventanas de Opciones del editor. Para terminar, se selecciona la opción Add Rows (Añadir filas) en el menú-ventana de Opciones del editor (véase la Figura 3-41).

- Usa la Flecha hacia abajo para poner el cursor en la celda 11 (valor de datos .211) en la variable sales.
- Pulsa F5 (Opts).
- Usa la Flecha hacia abajo para seleccionar la opción Add Rows (Añadir filas).

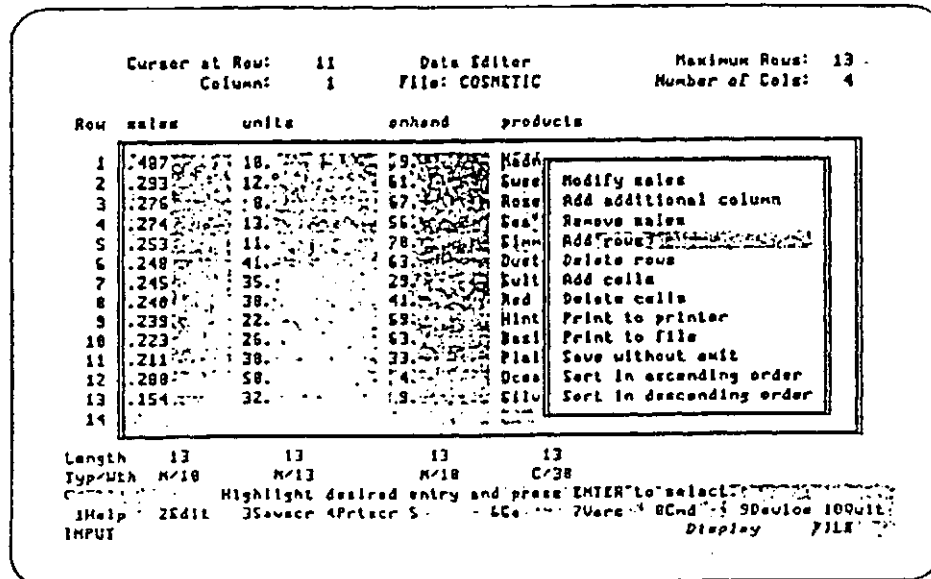


Figura 3-41. Menú-ventana de Opciones del editor con selección de la opción Add Rows (Añadir filas)

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar la selección hecha en el menú-ventana de Opciones del editor, el sistema pide que se especifique el número de filas a añadir (véase la Figura 3-42). Hay que añadir **una** fila a las variables **sales**, **units**, **onhand**, y **products**, así que pulsa la tecla Enter para aceptar el valor de la opción por defecto.

- Pulsa la tecla Enter otra vez para aceptar el valor de una fila a añadir como opción por defecto.



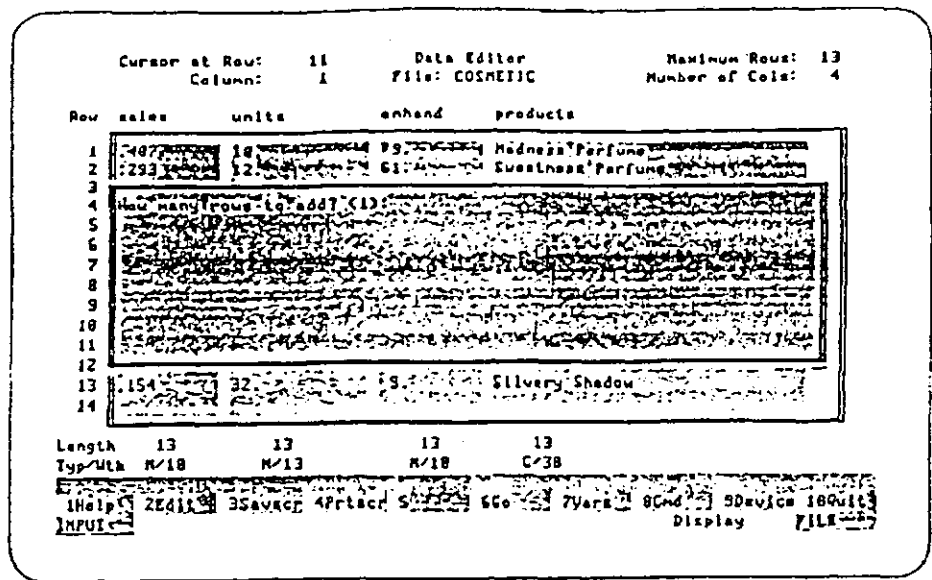


Figura 3-42. Indicador de número de filas a añadir

Una vez que se pulsa la tecla Enter, el sistema añade una fila en blanco en la fila 11 y mueve las siguientes un puesto hacia abajo (véase la Figura 3-43). (Cuando se usa la opción para Añadir filas, STATGRAPHICS empieza a añadir filas en la fila en la que se encuentre el cursor).

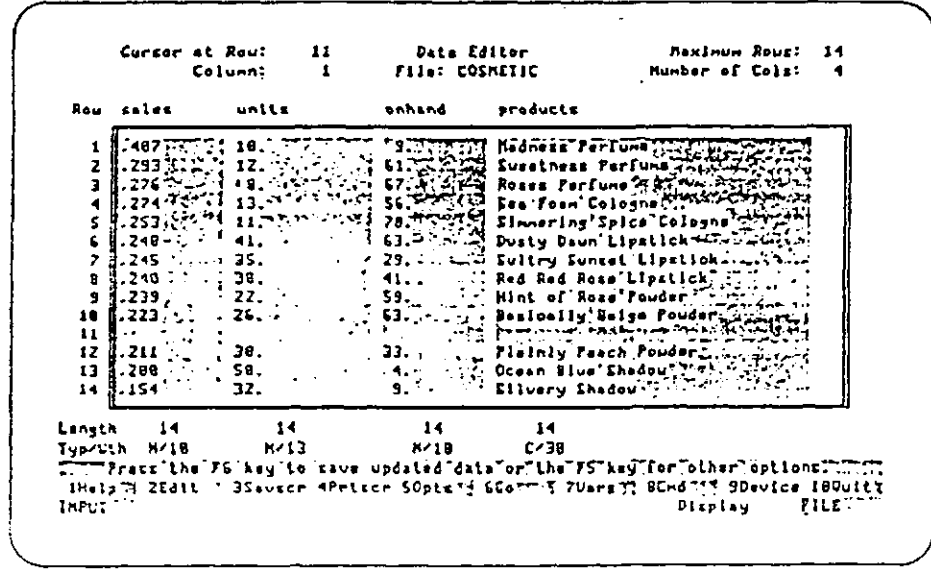


Figura 3-43. Variables con una fila en blanco añadida en el editor de datos

Ahora se puede poner datos en la Fila 11 para todas las variables. Se puede poner números enteros en las celdas en blanco de una variable numérica sin escribir el punto separador de decimales al final del valor (véase la Fig. 3-44). La próxima vez que se presenten las variables en el editor de datos, STATGRAPHICS añadirá los separadores de decimales y alineará los valores de los datos.

- Escribe el número .156 en la celda en blanco de la variable **sales**.
- Usa la tecla Tab para poner el cursor en la variable **units**. Escribe el número 37 en la celda en blanco.
- Usa la tecla Tab para poner el cursor en la variable **onhand**. Escribe el número 12 en la celda en blanco.
- Usa la tecla Tab para poner el cursor en la variable **products**. Escribe el nombre Gentle Grey Shadow en la celda en blanco.

Cursor at Row: 11      Data Editor      Maximum Row: 14  
                  Column: 4      File: COSMETIC      Number of Col: 4

Row	sales	units	onhand	products
1	487	10	29	Madness Perfume
2	293	12	61	Sweetness Perfume
3	276	18	67	Roses Perfume
4	274	13	66	Sea Foam Cologne
5	253	11	78	Glimmering Spice Cologne
6	248	41	63	Bustly Dawn Lipstick
7	245	35	29	Exhilarating Lipstick
8	248	36	41	Red Red Rose Lipstick
9	239	22	59	Hint of Rose Powder
10	223	25	63	Basically Beige Powder
11	156	37	12	Gentle Grey Shadow
12	211	39	33	Plainly Peach Powder
13	288	58	14	Ocean Blue Shadow
14	154	32	19	Silvery Shadow

Length    14               14               14               14  
Typ/Uch   N/10           N/13               N/18              C/38

Press the F5 key to save updated data or the F5 key for other options.  
[Help] [2Edit] [3Save] [4Print] [5Options] [6Quit] [7Yours] [8Cancel] [9Delete] [10Quit]  
Input: \_\_\_\_\_ Display: \_\_\_\_\_

Figura 3-44. Variables con datos en la nueva fila en el editor de datos

La columna de **products** es más ancha de lo necesario para los nombres de productos que hay en ella. Vamos a usar el menú-ventana de Opciones del editor para modificar la amplitud de la columna de **products**. El primer paso es pulsar F5 (Opts) para abrir el menú-ventana de Opciones del editor, y luego seleccionar la opción **Modify** para modificar **products** (véase la Figura 3-45).

- Pulsa F5 (Opts).
- Usa la tecla Home, si es necesario, para seleccionar la opción Modify (Modificar) products.

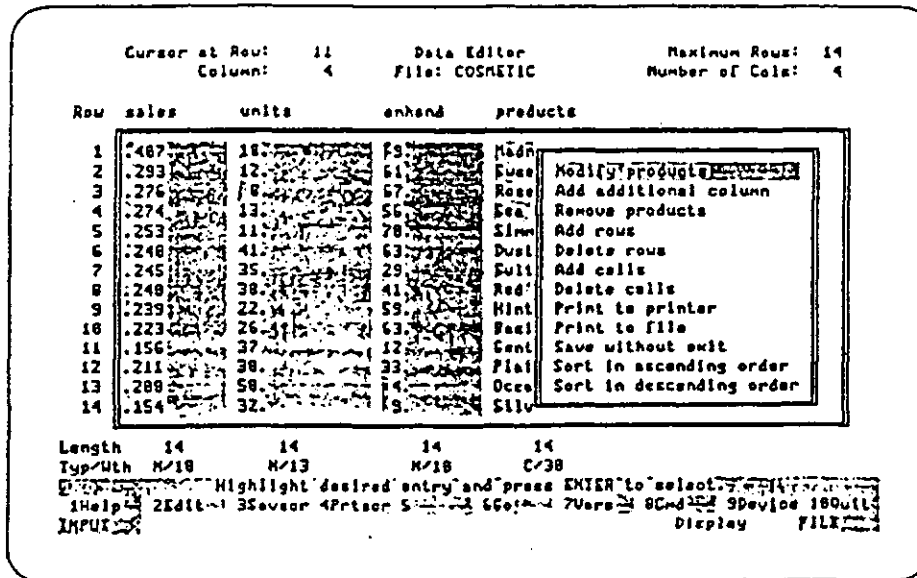


Figura 3-45. Menú-ventana de Opciones del editor con selección de la opción Modify (Modificar) products

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar la selección hecha en el menú-ventana de Opciones del editor, el sistema presenta la ventana Modify Current Column (Modificar la columna activada). Se usa la ventana para Modificar la columna activada al objeto de indicar la nueva amplitud (véase la Figura 3-46).

- Usa la tecla Tab para ir al campo Width (Amplitud).
- Escribe el número 25 encima del número 30 en este campo.

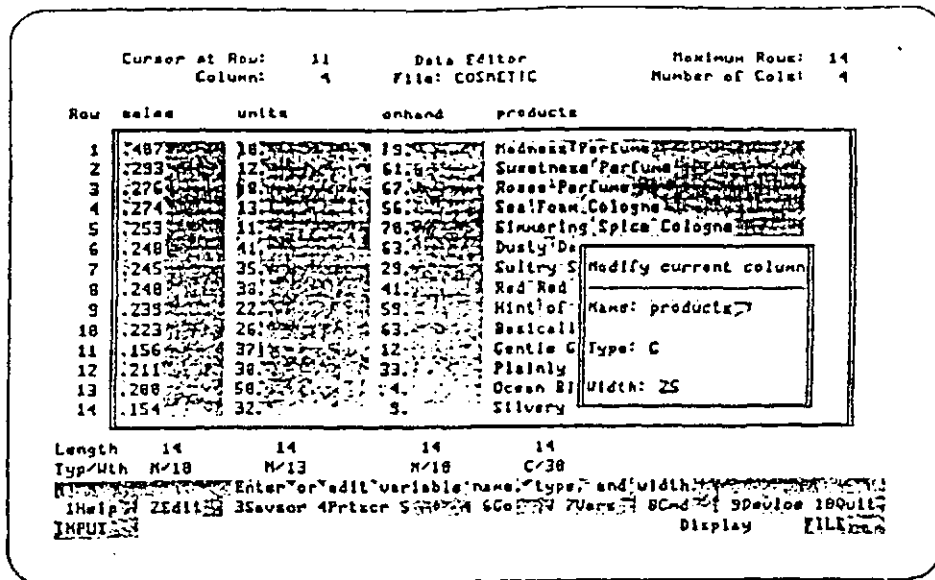


Figura 3-46. Ventana para Modificar la columna activada con indicación de la nueva amplitud

- Pulsa F6 (Go)

Una vez que se pulsa F6 (Go) para procesar la ventana para Modificar la columna activada, el sistema cambia la amplitud de la columna y modifica la anotación Typ/Wth (Tipo/Amplitud) al final de la columna de products para reflejar el cambio de amplitud (véase la Figura 3-47).

- Pulsa F6 (Go)

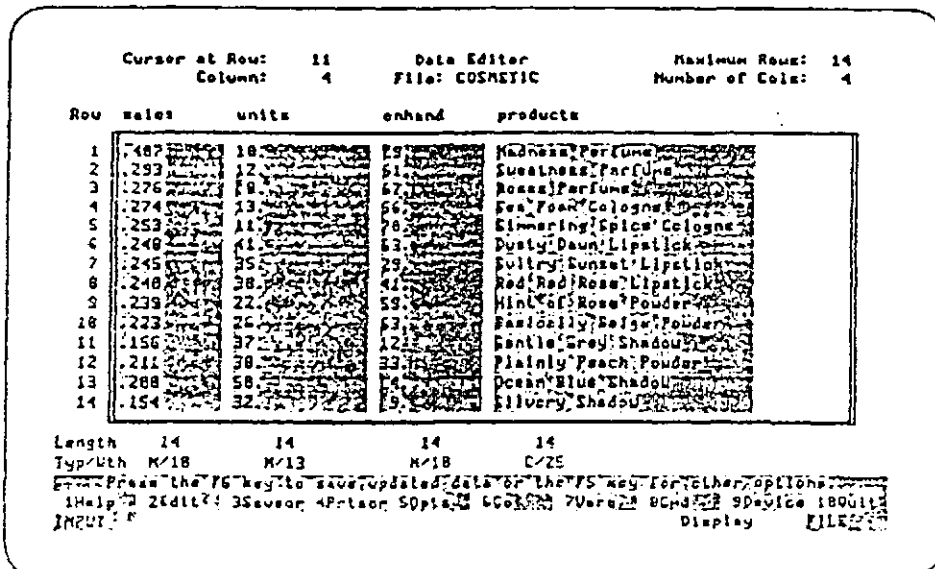


Figura 3-47. Columna modificada de la variable products en el editor de datos

Una vez que se pulsa otra vez F6 (Go) para procesar las entradas en la ventana para Modificar la columna activada, el sistema presenta el menú-ventana de Almacenaje. Hay que almacenar los datos y salir del editor de datos. Selecciona la opción para Almacenar y Salir en el menú-ventana para Almacenar (véase la Figura 3-48).

- Usa la Flecha hacia abajo para seleccionar la opción Save and Exit (Almacenar y Salir).

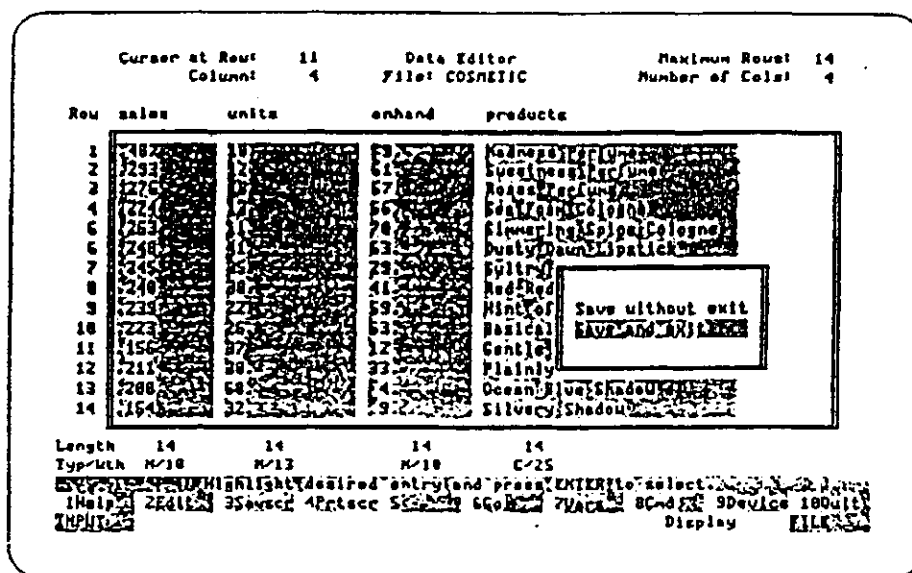


Figura 3-48. Menú-ventana Save (Almacenar) con selección de la opción Save and Exit (Almacenar y Salir)

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar la selección hecha en el menú-ventana para Almacenar, el sistema almacena los cambios, sale del editor de datos, y pasa a la pantalla de Contenido de ficheros (véase la Figura 3-49).

The following variables are currently in the file COSMETIC:

Variable	Width	Type	Rank	Length	Date	Time	Comment
sales	18	N	1	14	11/12/98	13:05	
units	13	N	1	14	11/12/98	13:05	
anhand	18	N	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	

Press ENTER key for each variable to be edited, or press R=All Display, H=Help, 2=Edit, 3=Save, 4=Print, 5=Quit, 6=Go, 7=Warn, 8=End, 9=Delete, 10=Quit, INPUT, Display, FILE

Figura 3-49. Pantalla de Contenido del fichero COSMETIC

Una vez terminada esta lección, vuelve al Menú principal. Se vuelve al Menú principal saliendo del procedimiento de Operaciones con ficheros (véase la Fig.3-50)

- Pulsa la tecla Esc hasta volver al Menú principal.

STATGRAPHICS Statistical Graphics System	
<b>DATA MANAGEMENT AND SYSTEM UTILITIES</b> A. Data Management B. System Environment C. Report Writer and Graphics Replay D. Graphics Attributes	<b>TIME SERIES PROCEDURES</b> L. Forecasting M. Quality Control N. Smoothing O. Time Series Analysis
<b>PLOTTING AND DESCRIPTIVE STATISTICS</b> E. Plotting Functions F. Descriptive Methods G. Estimation and Testing H. Distribution Functions I. Exploratory Data Analysis	<b>ADVANCED PROCEDURES</b> P. Categorical Data Analysis Q. Multivariate Methods R. Nonparametric Methods S. Sampling T. Experimental Design
<b>ANOVA AND REGRESSION ANALYSIS</b> J. Analysis of Variance K. Regression Analysis	<b>MATHEMATICAL AND USER PROCEDURES</b> U. Mathematical Functions V. Macros and User Functions
Use cursor keys to highlight desired function. Then press ENTER. H=Help, 2=Edit, 3=Save, 4=Print, 5=Quit, 6=Go, 7=Warn, 8=End, 9=Delete, 10=Quit, INPUT, Display, FILE	

Figura 3-50. El Menú principal

## RESUMEN

En este capítulo trabajaste con variables haciendo uso del editor de datos en el procedimiento de Operaciones con ficheros.

Concretamente, en este capítulo :

- hiciste selecciones en menú para llegar al procedimiento de Operaciones con ficheros.
- usaste la pantalla de entrada de datos para las Operaciones con ficheros y la opción C (Editar) para crear un fichero nuevo y entrar en el editor de datos.
- usaste el editor de datos para crear variables y poner los valores de los datos.
- usaste la opción D (Visualización) y la opción A (Todas) en la pantalla de Contenido de ficheros para ver los contenidos de las variables.
- usaste **F5 (Opts)** (Opciones) para crear una nueva variable y poner valores de datos para dicha variable.
- usaste el editor de datos para modificar los valores de los datos, borrar una fila, y añadir una fila a las cuatro variables.
- usaste **F5 (Opts)** (Opciones) para modificar la amplitud de una variable.
- usaste la opción Save and Exit (Almacenar y Salir) en el menú-ventana Save para almacenar los cambios y salir del editor de datos.

## MANEJO DE DATOS CON LA OPERACION DE ACTUALIZACION.

En el Capítulo 3 aprendiste cómo usar el editor de datos para crear ficheros y variables. En este capítulo vas a aprender a usar la operación de Actualización del procedimiento de Operaciones de fichero, al objeto de crear y editar (modificar) las variables.

Los ejercicios de este capítulo emplean el fichero COSMETIC que creaste en el Capítulo 3. Este fichero contiene variables con estadísticas de ventas (millones de dólares), unidades vendidas (millares), e inventario disponible (millares), de 14 productos cosméticos de Worldwide Products.

En este capítulo aprenderás a crear variables y a editarlas con la operación de Actualización del procedimiento de Operaciones de fichero. También aprenderás a emplear los operadores (símbolos y palabras especiales que permiten manejar y transformar datos) de STATGRAPHICS.

## USO DE DOS VARIABLES NUMERICAS PARA CALCULAR UNA TERCERA.

En esta lección se usa la ventana de Asignación de la operación de Actualización para calcular una variable numérica nueva (**price**) usando dos variables existentes (**sales y units**). También se hace uso de los siguientes operadores de STATGRAPHICS : ROUND, \* (multiplicación), y / (división).

La Lección 4-1 empieza en el Menú principal.

El primer paso para crear una variable es emplear los menús para seleccionar el procedimiento de Operaciones de fichero. El procedimiento de Operaciones de fichero está en el menú secundario de Gestión de datos, así que empieza por seleccionar el menú secundario de Gestión de datos en el Menú principal (véase la Figura 4-1).

- Si hace falta, usa la tecla Home para seleccionar el menú secundario de Gestión de datos.

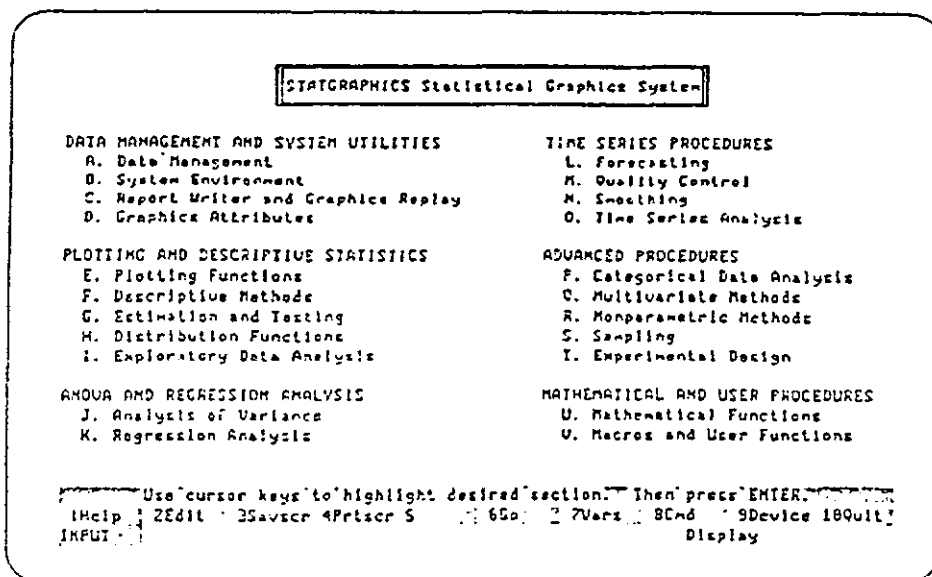


Figura 4-1. Menú principal con selección del menú secundario de Gestión de datos



- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo seleccionado en el Menú principal, el sistema presenta el menú secundario de Gestión de datos. Selecciona el procedimiento de Operaciones de fichero desde el menú secundario de Gestión de datos (véase la Figura 4-2).

- Usa la Flecha hacia abajo para seleccionar el procedimiento de Operaciones de fichero.

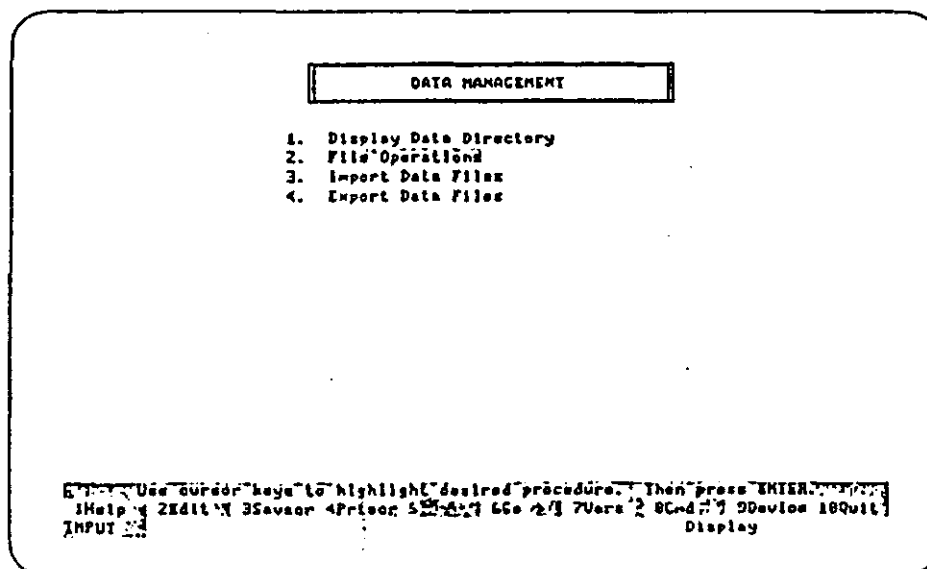


Figura 4-2. Menú secundario de Gestión de datos con selección del procedimiento de Operaciones de fichero

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo seleccionado en el menú secundario de Gestión de datos, el sistema presenta la pantalla de entrada de datos para Operaciones de fichero (véase la Figura 4-3). Se usa esta pantalla para indicar el fichero de datos y la operación con los que trabajar.

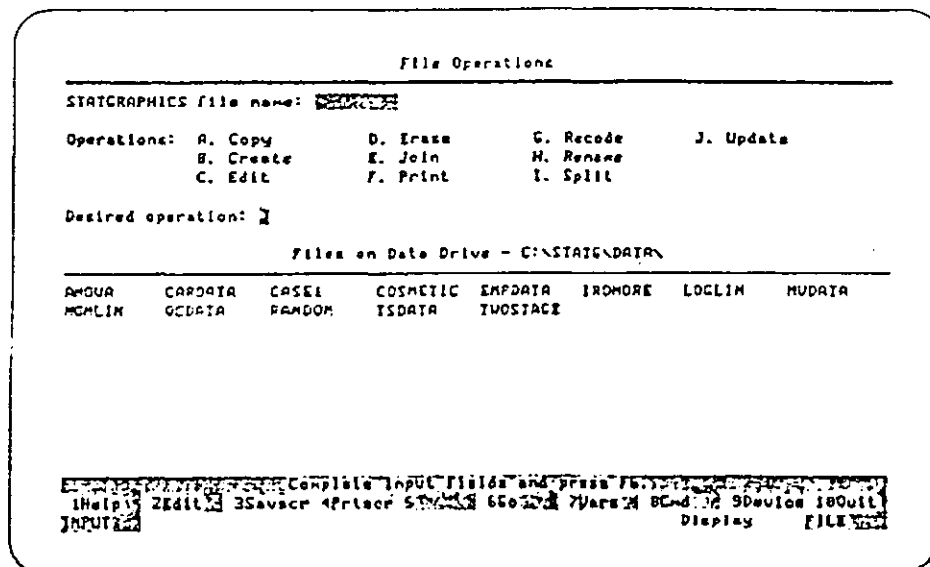
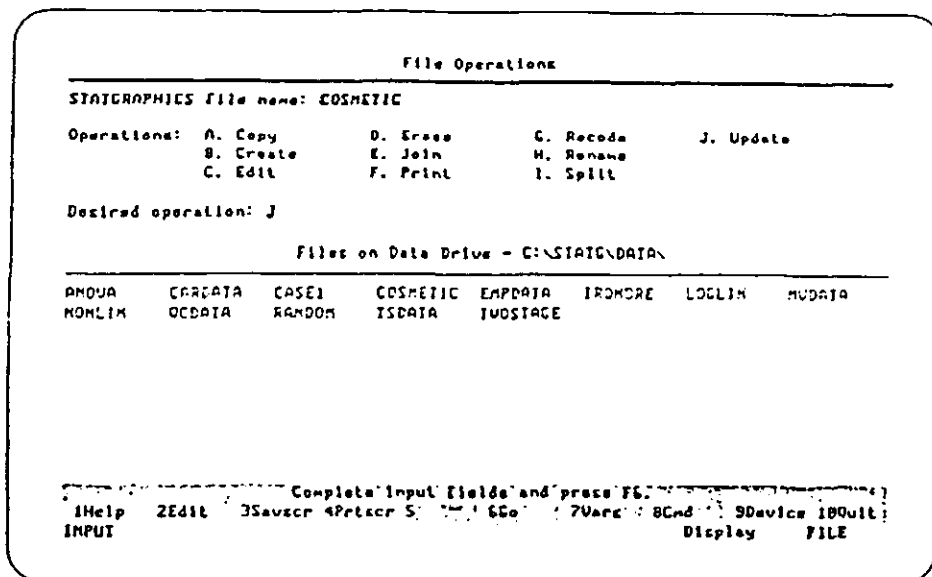


Figura 4-3. Pantalla de entrada de datos para Operaciones de fichero

Hay que decirles a STATGRAPHICS el nombre del fichero con el que trabajar (COSMETIC) e indicar la operación (Actualización) deseada (véase la figura 4-4).

- Escribe el nombre del fichero **COSMETIC** en el campo **STATGRAPHICS File Name**. El cursor va automáticamente al campo **Desired Operation** (Operación deseada).
- Si es necesario, escribe la letra **J** (Updtate- Actualizar) en este campo.



File 4-4. Pantalla de entrada de datos para Operaciones de fichero una vez completada

- Pulsa **F6 (Go)**.

Una vez que se pulsa **F6 (Go)** para procesar las entradas en la pantalla de entrada de datos para Operaciones de fichero, el sistema presenta la pantalla de Actualización de variables, con el encabezamiento "Contents of file COSMETIC" (Contenido del fichero COSMETIC) (véase la Figura 4-5). Esta pantalla presenta una lista de las variables del fichero con el que se esté trabajando.

Queremos crear una variable nueva. Mira la Línea de estado. La opción N (New-Nueva) en la pantalla de Actualización de variables nos permite crear una variable nueva en el fichero COSMETIC.

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 8/98	16:28	
price	16	N	1	14	11/ 8/98	17:52	
units	13	N	1	14	11/ 8/98	17:52	
onhand	18	N	1	14	11/ 8/98	17:52	
products	25	C	2	14 25	11/ 8/98	17:52	

[Highlight] [O]fiable [A]sAssign [C]omment [D]isplay [E]x[ist] [M]enu [R]ename [V]Copy  
 [I]nput [Z]edit [S]ave [P]rint [S]ort [C]om [Y]es [C]ancel [D]elete [O]ut  
 INPUT

Figura 4-5. Pantalla de Actualización de variables

- Escribe la letra N (Nueva) en este campo.

Una vez que se escribe la letra N para añadir una variable nueva, el sistema presenta el campo Variable nueva. Hay que crear una variable nueva, **price**, para que contenga el precio medio de cada uno de los 14 productos cosméticos cuyas estadísticas se encuentran en el fichero COSMETIC. Escribe **price** en el campo Variable nueva para decirle a STATGRAPHICS en nombre de la variable nueva (véase la Figura 4-6).

- Escribe el nombre price en el campo.

Contents of file COSMETIC							
Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 8/98	16:28	
sales	14	N	1	14	11/ 8/98	17:52	
units	13	N	1	14	11/ 8/98	17:52	
onhand	18	N	1	14	11/ 8/98	17:52	
products	25	C	2	14 25	11/ 8/98	17:52	

Enter name of new variable: price

[H]ighlight [V]ariable [A]ssign [C]omment [D]isplay [E]rase [N]ew [R]ename [V]copy  
 [H]elp [E]dit [S]ave [P]rint [C]opy [V]ar [D]evice [O]utput  
 INPUT Display FILE

Figura 4-6. Campo New Variable (Variable nueva) una vez completado

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar la Variable nueva, el sistema presenta la ventana de Asignación. Se usa la ventana de Asignación para indicar la fórmula (ecuación) que habrá de crear los valores de la variable price. Pero, ¿Cuál es la fórmula apropiada de la variable price ?

Sabemos que la variable sales del fichero COSMETIC contiene ventas en millones de dólares. Por ejemplo, Madness Perfume tiene 0,407 millones de dólares de ventas- \$ 407.000. También sabemos que la variable units del fichero COSMETIC contiene el número de unidades vendidas en millares de unidades. Por ejemplo, Madness Perfume ha vendido 10 millares de unidades - 10.000 unidades. Para calcular el precio medio de Madness Perfume hay que multiplicar sales por 1000 (para convertir 0,407 a miles de dólares), y dividir el resultado por las unidades (en millares). Haciendo uso del operador \* (multiplicación) y del operador / (división), la fórmula es :

$$(0,407 * 1000) / 10 =$$

$$407 / 10 =$$

$$40,7$$

Es decir, el precio medio de Madness Perfume es \$ 40,70.

Podemos usar la misma fórmula para calcular el precio medio de todos los productos en el fichero COSMETIC. Todo lo que hay que hacer es poner en la fórmula los nombres de las variables apropiadas :

$$(\text{sales} * 1000) / \text{S}$$

También tenemos que redondear el precio medio de cada producto convirtiéndolo al número íntegro más cercano. Para redondear un valor al número entero más cercano se usa el operador ROUND. (Observa que al escribir el nombre de un operador **hay que hacerlo** en letras mayúsculas). La fórmula para calcular el precio medio de Madness Perfume es :

$$(.407 * 1000) / 10$$

Para redondear el precio medio al número íntegro más cercano, la fórmula es:

$$\text{ROUND} ((.407 * 1000) / 10) =$$
$$\text{ROUND} (407 / 10) =$$
$$\text{ROUND} 40.7 =$$

41

Observa que hay que poner entre paréntesis la parte de la fórmula para calcular el precio medio al objeto de que el operador **ROUND** redondee el resultado del cálculo completo del promedio.

Como antes, se pueden poner los nombres de las variables en la fórmula para redondear los precios medios de todos los productos. La fórmula para crear los valores de la variable price (véase la Figura 4-7) es :

$$\text{ROUND} (\text{sales} * 1000) / \text{units}$$

Nota: Si todos los nombres de las variables en el directorio de datos son diferentes, se puede poner solo el nombre de la variable. Pero hay dos variables llamadas **sales** y dos variables llamadas **units** en el directorio de datos- una en el fichero COSMETIC y otra en el fichero llamado TSDATA. Por ello hay que poner en la fórmula el nombre del fichero y el nombre de la variable de esta manera:

fichero.variable

Por ejemplo, se usa

`COSMETIC.sales`

para referirse a la variable `sales` en el fichero `COSMETIC`.

- En la ventana de Asignación, escribe la fórmula **exactamente** como se indica a continuación:

```
ROUND (( COSMETIC .sales * 1000) /  
COSMETIC . units)
```

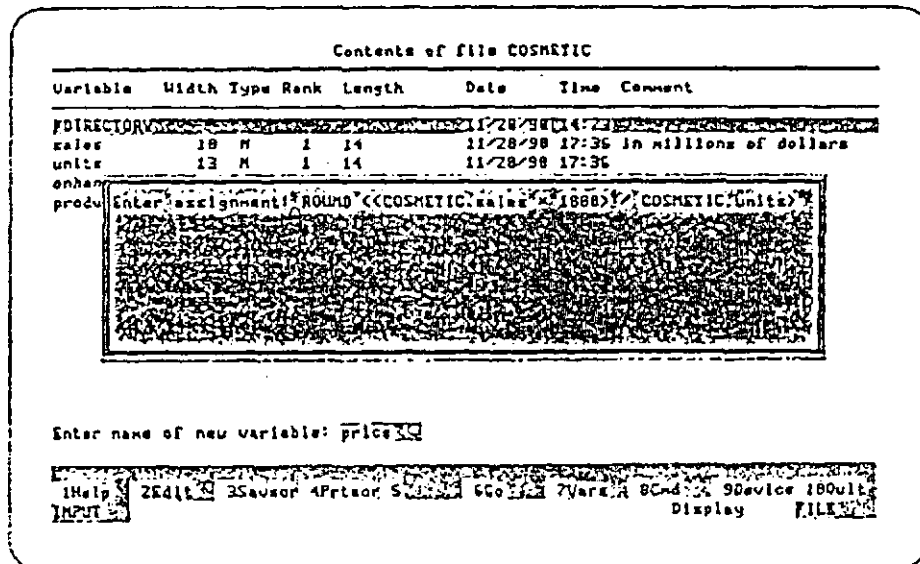


Figura 4-7. Ventana de Asignación con la fórmula para la variable `price`

- Pulsa la tecla `Enter`.

Una vez que se pulsa la tecla `Enter` para procesar lo entrado en la ventana de Asignación, el sistema presenta el campo `Comment` (Comentario). Este campo se usa para poner una descripción breve de la nueva variable (véase la Figura 4-8).

Escribe el comentario precio medio en el campo.

Contents of file COSMETIC						
Variable	Width	Type	Rank	Length	Date	Time Comment
FDIRECTORY					11/ 9/98	16:28
sales	18	N	1	14	11/12/98	13:05
units	13	N	1	14	11/12/98	13:05
onhand	18	N	1	14	11/12/98	13:05
products	25	C	2	14 25	11/12/98	13:05

Enter comment: average price

1Help 2Edit 3Save 4Print 5 6Go 7Vars 8Cmd 9Device 10Quit  
INPUT Display FILE

Figura 4-8. El campo Comment (Comentario) una vez completado

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo escrito en el campo Comment, el sistema almacena los valores de datos en la nueva variable y almacena el comentario en el campo Comment. El sistema añade la variable price al final de la lista de variables en la pantalla de Actualización de variables (véase la Fig. 4-9).

Ahora puedes pasar a la Lección 4-2. En la Lección 4-2, vas a aprender cómo actualizar variables existentes. La Lección 4-2 empieza en la pantalla de Actualización de variables



Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY	10	N	1	14	11/9/98	16:20	
sales	10	N	1	14	11/12/98	13:05	
units	13	N	1	14	11/12/98	13:05	
onhand	10	N	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
price		N	1	14	11/12/98	15:48	average price

[Highlight Variable: A=Assign C=Comment D=Display E=Erase M=New R=Rename V=Copy]  
 [Help] [Edit] [3Save] [4Print] [5Search] [6Go] [7Vars] [8Cmd] [9Dev] [10Quit]  
 INPUT \_ \_ \_ Display FILE

Figura 4-9. Pantalla de Actualización de variables con la variable price

### EDICION DE VARIABLES EXISTENTES.

La operación de Actualización ofrece varias opciones para actualizar ficheros además de las aprendidas- Asignación y Nueva. Esta lección enseña cómo usar las opciones restantes, indicadas más abajo, en variables existentes. Las variables se pueden editar de una en una.

- Display- Visualizar
- Comment- Comentar
- Rename- Cambiar de nombre
- Copy- Copiar
- Erase- Borrar.

La Lección 4-2 empieza en la pantalla de Actualización de la variable.

Queremos examinar el contenido de la variable price. Primero, selecciona la variable price (véase la Figura 4-10). Luego usa la opción D (Display-Visualizar) para presentar el contenido de la variable.

- Use la Flecha hacia abajo para seleccionar la variable price.

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 8/98	16:28	
sales	18	M	1	14	11/ 8/98	17:52	
units	13	M	1	14	11/ 8/98	17:52	
enhend	10	M	1	14	11/ 8/98	17:52	
products	25	C	2	14 25	11/ 8/98	17:52	
price	14	M	1	14	11/ 8/98	18:11	average price

[Highlight variable: ~A=Assign C=Comment D=Display E=Erase M=Menu R=Rename Y=Copy]  
 [Help 1 ZEdit 3 Ssave 4Prscr 5 : 6Go 7Vars 8Cmd 9Device 10Quit]  
 INPUT: Display FILE:51

Figura 4-10. Pantalla de Actualización de variables con la variable price seleccionada

- Escribe la letra D (Display-Visualizar).

Una vez que escribas la letra D para ver los valores de datos de la variable price el sistema presenta la pantalla de Contenido (véase la Figura 4-11). La pantalla de Contenido muestra el nombre de la variable, su longitud y los valores de los datos. Desde esta pantalla no se pueden editar los valores de los datos almacenados en la variable.

Variable: COSMETIC.price (length = 14)

< 1> 41
< 2> 24
< 3> 35
< 4> 21
< 5> 22
< 6> 6
< 7> 7
< 8> 6
< 9> 18
< 10> 8
< 11> 4
< 12> 7
< 13> 4
< 14> 5

Press Esc, Cursor keys or Page Number to move cursor. Page 1 of 1  
 [Help 1 ZEdit 3 Ssave 4Prscr 5 Ptopt 6Go 7Vars 8Cmd 9Device 10Quit]  
 INPUT: Display FILE:51

Figura 4-11. Pantalla de Contenido de la variable price

Ahora vamos a usar la opción C (Comment-Comentar) para hacer un comentario sobre la variable sales. Usa la tecla Esc para volver a la pantalla de Actualización de variables. Luego, selecciona la variable sales en la pantalla de Actualización de variables (véase la Figura 4-12) y especifica la opción C (Comment-Comentario).

- Pulsa la tecla Esc.
- Usa la Flecha hacia arriba para seleccionar sales.

Contents of file COSMETIC							
Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 8/98	16:28	
Sales	18	N	1	14	11/ 8/98	17:52	
units	13	N	1	14	11/ 8/98	17:52	
onhand	18	N	1	14	11/ 8/98	17:52	
products	25	C	2	14 25	11/ 8/98	17:52	
price		N	1	14	11/ 8/98	18:11	average price

[H]ighlight [U]nassign [A]sAssign [C]omment [D]isplay [E]rase [M]enu [R]ename [V]Copy  
 [M]help [Z]fill [S]earch [P]rint [S]ort [G]o [C]opy [V]ars [B]ack [S]ave [D]evice [Q]uit  
 INPUT: Display FILE:

Figura 4-12. Pantalla de Actualización de variables con la variable sales seleccionada

- Escribe la letra C (Comentario).

Una vez que escribes la letra C para añadir un comentario a la variable sales variable, el sistema presenta el campo Comment. Escribe a descripción in millions of dollars en la variable sales (véase la Figura 4-13).

- Escribe el comentario en "millones de dólares" en el campo.

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 9/98	16:28	
sales	12	N	1	14	11/12/98	12:05	
units	13	N	1	14	11/12/98	13:05	
onhand	10	N	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
price		N	1	14	11/12/98	15:48	average price

Enter comment: in millions of dollars

[H]ighlight [V]ariable [A]ssign [C]omment [D]isplay [E]rase [N]ew [R]ename [Y]Copy  
 [H]elp [Z]Edit [S]ave [P]rint [S]ort [G]o [U]n [C]md [D]evice [O]ut  
 INPUT - [ ] Display FILE

Figura 4-13. El campo Comment (Comentario) una vez completado

- pulsa la tecla Enter.

Al pulsar la tecla Enter, el sistema muestra la pantalla de Actualización de variables. Observa la nueva descripción en el campo Comment de la variable sales (Véase la Figura 4-14).

Ahora hay que usar la opción R (Rename-Cambiar de nombre) para cambiar el nombre de la variable price para que la describa mejor. Se selecciona la variable price en la pantalla de Actualización de variables (véase la Figura 4-14). Luego se usa la opción R (Rename) para cambiar el nombre de la variable.

- Usa la Flecha hacia abajo para seleccionar la variable price en esta pantalla.

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 9/98	16:28	
sales	18	M	1	14	11/12/98	13:05	In millions of dollars
units	13	M	1	14	11/12/98	13:05	
onhand	18	M	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
price	18	M	1	14	11/12/98	15:48	Average price

[Highlight variable] [A=Assign] [C=Comment] [D=Display] [E=Erase] [N=New] [R=Rename] [V=Copy]  
 [H=Help] [2=Edit] [3=Save] [4=Print] [5=...] [6=Go] [7=Vars] [8=...] [9=Dev] [10=Quit]  
 INPUT [ ] Display FILE [ ]

Figura 4-14. Pantalla de Actualización de variables con selección de la variable price

- Escribe la letra R (Rename- Cambiar el nombre).

Una vez que se escribe la letra R para cambiar el nombre de la variable, el sistema presenta el campo New Name (Nombre nuevo). Usa este campo para escribir el nombre nuevo de la variable `price_ avgprice` (véase la Figura 4-15).

\_ Escribe el nombre `avgprice` en el campo New Name.

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
PDIRECTORY					11/ 9/98	16:28	
sales	18	M	1	14	11/12/98	13:05	in millions of dollars
units	13	M	1	14	11/12/98	13:05	
onhand	18	M	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
<del>price</del>							

Enter new name: supplier

HIGHLIGHT	VARIABLES	AAssign	CComment	DDisplay	EExit	MMenu	RRename	VCopy
1Help	2Quit	3Save	4Print	5Search	6Go	7Var	8Dev	9Quit
INPUT							Display	[(LR)]

Figura 4-15. El campo New Name (Nombre nuevo) una vez completado

- Pulsa la tecla Enter.

A pulsar la tecla Enter para procesar lo escrito en el campo New Name, el sistema muestra la pantalla de Actualización de variables. Observa el nombre nuevo de la variable **price** en la pantalla de Actualización de variables (véase la Figura 4-16).

Ahora vamos a emplear la opción Copy para copiar la variable **onhand** a una variable nueva **onhand2**. Primero, selecciona la variable **onhand** en la pantalla de Actualización de variables (véase la Figura 4-16). Luego usa la opción Y (Copy-Copiar) para copiar la variable

- Usa la Flecha hacia arriba para seleccionar la variable **onhand**.

Contents of file COSMETIC							
Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 9/98	16:28	
sales	18	M	1	14	11/12/98	13:05	In millions of dollars
units	13	M	1	14	11/12/98	13:05	
onhand	18	M	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
avgprice		M	1	14	11/12/98	15:48	average price

[H]ighlight [V]ariable [A]ssign [C]omment [D]isplay [X]tend [N]ew [R]ename [Y]Copy  
 [I]Help [Z]Edit [3]Save [4]Print [5]Sort [6]Go [7]Vars [8]Cmd [9]Device [0]Quit  
 [N]PUT [2] Display [I]LK [3]

Figura 4-16. Pantalla de Actualización de variable con selección de la variable onhand

- Escribe la letra Y (Copy-Copiar).

Una vez que se escribe la letra Y para copiar la variable **onhand**, el sistema muestra el campo New Variable (Variable nueva). Este campo se usa para escribir el nombre de la variable nueva- **onhand2** (véase la Figura 4-17).

- Escribe el nombre onhand2 en el campo New.

Contents of file COSMETIC							
Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 9/98	16:28	
sales	18	M	1	14	11/12/98	13:05	In millions of dollars
units	13	M	1	14	11/12/98	13:05	
onhand	18	M	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
avgprice		M	1	14	11/12/98	15:48	average price

Enter name of new variable: onhand2

[Highlight] Variable: A=Assign C=Comment D=Display E=Erase M=New R=Rename Y=Copy  
 1Help 2Edit 3Save 4Print 5Quit 6Copy 7Vars 8Cmd 9Device 10Quit  
 INPUT Display FILE

Figura 4-17. El campo New Variable (Variable nueva) una vez completado

- Pulsa la tecla Enter.

Al pulsar la tecla Enter para procesar los escritos en el campo New Variable, el sistema presenta el campo Comment (Comentario) (véase la Figura 4-18).

Contents of file COSMETIC							
Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 9/98	16:28	
sales	18	M	1	14	11/12/98	13:05	In millions of dollars
units	13	M	1	14	11/12/98	13:05	
onhand	18	M	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
avgprice		M	1	14	11/12/98	15:48	average price

Enter comment:

[Highlight] Variable: A=Assign C=Comment D=Display E=Erase M=New R=Rename Y=Copy  
 1Help 2Edit 3Save 4Print 5Quit 6Copy 7Vars 8Cmd 9Device 10Quit  
 INPUT Display FILE

Figura 4-18. El campo Comment (Comentario) para la variable onhand2



No vamos a poner ningún comentario, así que pulsa la tecla Enter sin escribir comentario alguno para dejar el campo Comment de la variable **onhand2** en blanco (véase la Figura 4-19).

Como no se necesita la variable **onhand2** para cálculos futuros, vamos a borrarla. La opción E (Erase-Borrar) se usa para borrar la variable **onhand2**. Primero, selecciona la variable **onhand2** en la pantalla de Actualización de variables (véase la Figura 4-19). Luego, usa la opción E (Erase-Borrar) para borrar la variable.

- Pulsa la tecla Enter sin escribir ningún comentario.
- Usa la Flecha hacia abajo para seleccionar la variable **onhand2**.

Contents of file COSMETIC							
Variable	Width	Type	Rank	Length	Date	Time	Comment
<b>FDIRECTORY</b>					11/ 9/98	16:28	
sales	18	N	1	14	11/12/98	13:05	In millions of dollars
units	13	N	1	14	11/12/98	13:05	
onhand	18	N	1	14	11/12/98	13:05	
products	25	C	2	14 25	11/12/98	13:05	
avgprice		N	1	14	11/12/98	15:48	average price
<b>onhand2</b>	18	N	1	14	11/12/98	16:18	

!Highlight variable: A=Assign C=Comment D=Display E=Erase H=New R=Rename V=Copy  
 !Help 2Edit 3Sevscr 4Prscr 5 6Cc 7Uarc 8Cwd 9Device 10Quit  
 INPUT Display FILE

Figura 4-19. Pantalla de Actualización de variable con selección de la variable **onhand2**

- Escribe la letra E (Erase- Borrar).

Una vez que se escribe la letra E para especificar la opción de borrar, el sistema muestra la ventana de Borrar. En la ventana de Borrar, el sistema te pregunta si quieres o no borrar la variable **onhand2** (véase la Figura 4-20). En STATGRAPHICS esta opción es el método empleado para borrar una variable.

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 8/98	16:28	
sales	18	M	1	14	11/ 8/98	17:52	In millions of dollars
units	12	M	1	14	11/ 8/98	17:52	
onhand							
products							
avgpr							
onhand2							

OK to erase variable onhand2? (Y/N)

Help  ZEdit  3Saver  4Prtscr  5...  6Go  7Uars  8Cmd  9Device  18Quit  
 INPUT  Display  FILE

Figura 4-20. Ventana para borrar

- Escribe la letra Y (Yes - Sí).

Una vez que escribas Y (Yes - Sí) para indicar que deseas borrar la variable **onhand2**, el sistema quita la variable **onhand2** del fichero y de la lista de variables (véase la Figura 4-21)

Contents of file COSMETIC

Variable	Width	Type	Rank	Length	Date	Time	Comment
FDIRECTORY					11/ 8/98	16:28	
sales	18	M	1	14	11/ 8/98	17:52	In millions of dollars
units	12	M	1	14	11/ 8/98	17:52	
onhand	18	M	1	14	11/ 8/98	17:52	
products	25	C	2	14 25	11/ 8/98	17:52	
avgprice	14	F	1	14	11/ 8/98	18:11	average price

Highlight  Variables  A=Assign  C=Comment  D=Display  E=Erase  M=Mod  R=Rename  V=Copy  
 Help  ZEdit  3Saver  4Prtscr  5...  6Go  7Uars  8Cmd  9Device  18Quit  
 INPUT  Display  FILE

Figura 4-21. Pantalla de Actualización de variables con la variable **onhand2** borrada

Ahora que ya has terminado esta lección, puedes volver al Menú principal. Se vuelve al Menú principal saliendo del procedimiento de Operaciones de fichero (véase la Figura 4-22).

- Pulsa la tecla Esc hasta volver al Menú principal.

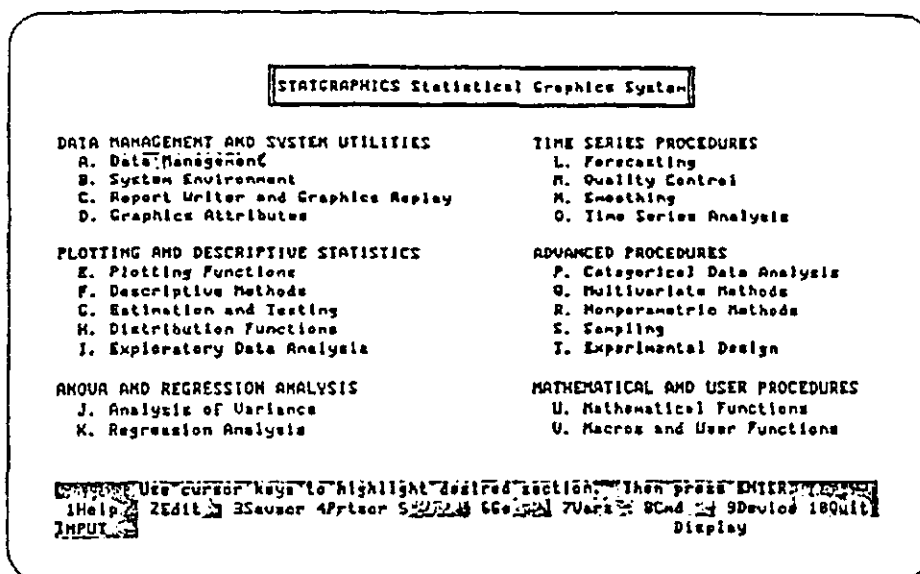


Figura 4-22. El Menú principal

## IMPORTACION DE DATOS.

A menudo, los datos a analizar vienen de otras fuentes. Puede que la sección de contabilidad te dé un disco con información sobre cuentas por cobrar producido por algún programa de gestión de base de datos. O que la sección de personal te dé un disco con datos sobre vacaciones de los empleados producido con un programa de hoja de cálculo. STATGRAPHICS te permite usar datos venidos de otras fuentes sin que haga falta volver a partir de cero.

Por medio del procedimiento de Importación de ficheros de datos en STATGRAPHICS, puedes emplear ficheros preparados con los formatos siguientes :

- ASCII (formados con delimitadores en blanco o con delimitadores de caracteres)
- Lotus 1-2-3, hasta la versión 2.2 (se puede importar un fichero de la versión 3.0 si antes se transforma a la versión 2.2 en 1-2-3)

- Symphony, hasta versión 1.2
- dBASE, hasta la versión IV
- ATLAS \* GRAPHICS
- DIF (Formato para intercambio de datos).

En este capítulo vas a aprender cómo importar un fichero llamado EMPDATA.DAT. Este fichero ASCII con delimitadores en blanco, contiene información del personal de Worldwide Products en la que se incluye el número de identidad, el nombre, fecha de empleo, salario a la hora, y horas semanales trabajadas. Decimos que este tipo de fichero tiene delimitadores en blanco porque se usan espacios en blanco para separar las columnas.

### **IMPORTACION DE UN FICHERO.**

En esta lección se usa el procedimiento de Importación de ficheros de datos para importar un fichero ASCII, con delimitadores en blanco, llamado EMPDATA.DAT. El fichero contiene información del personal de Worldwide Products en la que se incluye el número de identidad, el nombre, fecha de empleo, salario a la hora, y horas semanales trabajadas. Como ya tienes instalado STATGRAPHICS en el ordenador, el fichero EMPDATA.DAT que vino con STATGRAPHICS está ya almacenado y listo para su uso.

La Lección 5-1 empieza en el Menú principal.

El primer paso para importar el fichero EMPDATA.DAT es emplear los menús para seleccionar el procedimiento de Importación de ficheros de datos. El procedimiento de Importación de ficheros de datos se encuentra en el menú secundario de Gestión de datos, así que empieza por seleccionar el menú secundario de Gestión de datos en el Menú principal (véase la Figura 5-1).

- Usa la tecla Home, si hace falta, para seleccionar el menú secundario de Gestión de datos.

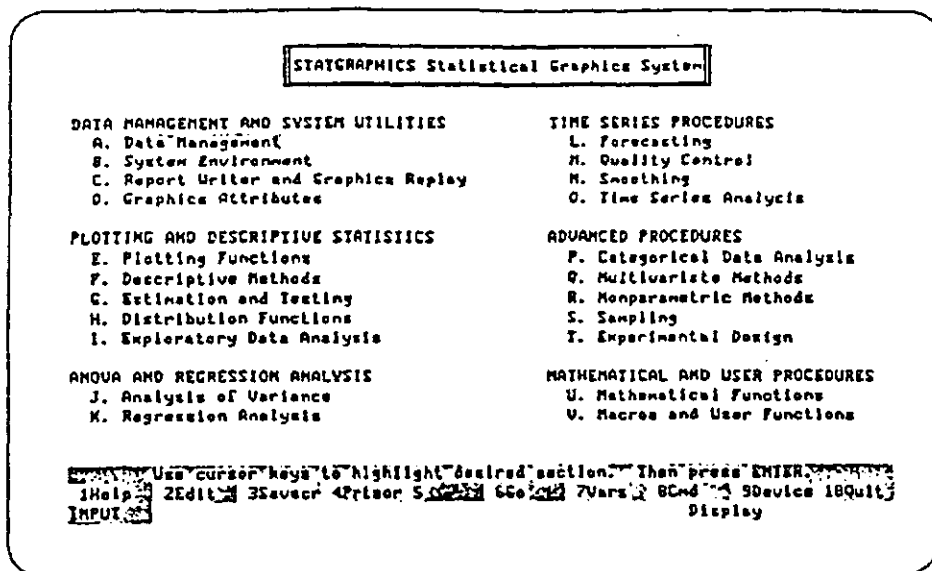


Figura 5-1. Menú principal con el menú secundario de Gestión de datos seleccionado

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo seleccionado en el Menú principal, el sistema presenta el menú secundario de Gestión de datos. Selecciona el procedimiento de Importación de ficheros de datos en el menú secundario de Gestión de datos (véase la Figura 5-2).

- Usa la Flecha hacia abajo para seleccionar el procedimiento de Importación de ficheros de datos.

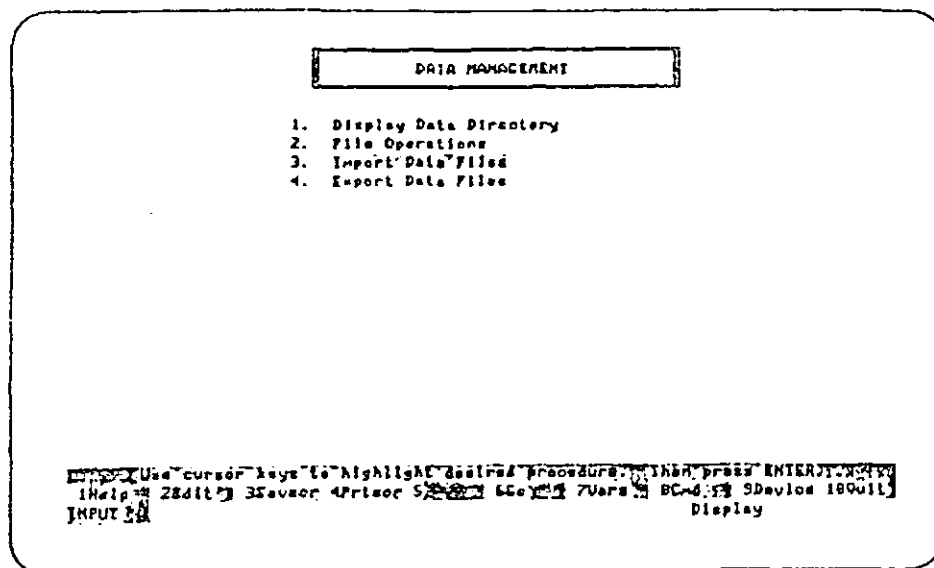


Figura 5-2. El menú secundario de Gestión de datos con el procedimiento de Importación de ficheros de datos seleccionado

- Pulsa la tecla Enter.

Una vez que se pulsa la tecla Enter para procesar lo seleccionado en el menú secundario de Gestión de datos, el sistema presenta la pantalla de entrada de datos para importación de ficheros (véase la Figura 5-3). Se emplea esta pantalla para darle al sistema detalles sobre el fichero EMPDATA.DATA que se va a importar.

Observa el centro de la Figura 5-3. Debajo de la línea que dice al principio "Files on Import Drive" (Ficheros a importar en el disco), el sistema presenta todos los ficheros de datos que no son ASF (que no son de STATGRAPHICS) en la ruta del directorio de importación (el directorio y subdirectorios que indican la ruta para llegar a la zona del disco donde se encuentran almacenados los ficheros). Los ficheros que aparecen en la lista son los ficheros que STATGRAPHICS considera disponibles para que los importes.

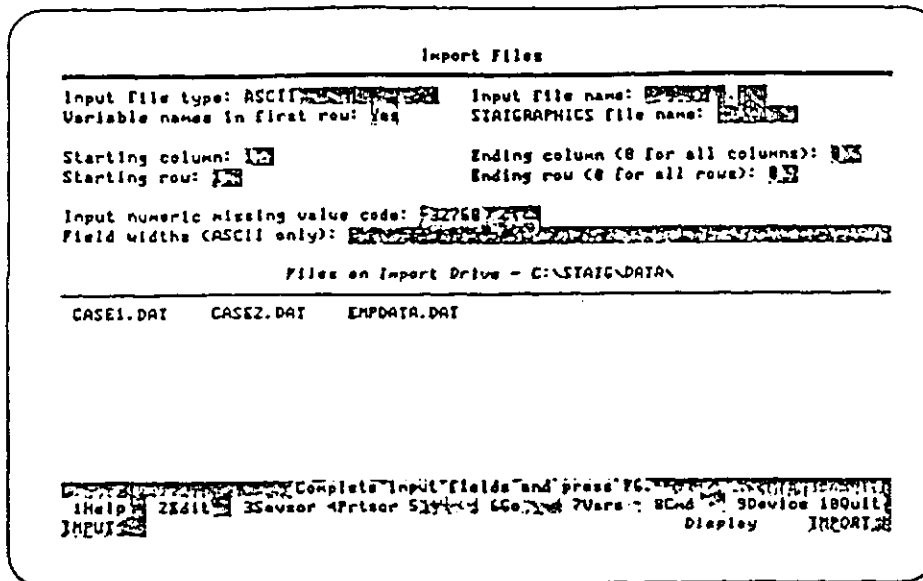


Figura 5-3. Pantalla de entrada de datos para Importación de ficheros

En la Figura 5-4, se muestra el contenido del fichero EMPDATA.DAT que se quiere importar. Observa que cada columna del fichero EMPDATA.DAT está separada por uno o más espacios en blanco. Decimos que este tipo de fichero tiene delimitadores en blanco porque hay espacios en blanco entre las columnas.

La primera fila del fichero tiene los encabezamientos de cada columna; por ejemplo, ID, Name, y Hire\_Date (identificación, nombre y fecha de empleo). Algunos de los encabezamientos de las columnas tienen barras inclinadas (Wage/hr y Hrs/Week- Salario/hora y Horas/semana). Aunque se puede usar en los ficheros ASCII, la barra inclinada es un carácter ilegal para su uso en el nombre de una variable de STATGRAPHICS. Por ello, cuando STATGRAPHICS importa el fichero EMPDATA.DAT, el sistema cambia los caracteres ilegales a la raya de subrayar.

ID	Name	Hire_Date	Wage/Hr	Hrs/Week
0000	Smith	02/23/80	5.50	30.0
0001	Jones	03/30/80	6.25	40.0
0002	Spink	07/01/80	4.75	20.0
0003	Fogel	09/30/80	4.55	25.0
0004	Whitson	01/01/81	3.95	37.5
0005	Wise	05/31/81	3.35	45.0
0006	Lindquist	10/31/81	4.00	36.0
0007	Baker	04/01/82	7.10	27.5
0008	Duerr	12/01/82	6.75	20.0
0009	White	06/30/83	5.50	20.0
0010	Montgomery	08/15/83	5.60	24.5
0011	Washington	02/29/84	10.20	25.5
0012	Battel	07/01/84	0.30	50.0
0013	Patterson	12/31/84	4.50	40.0
0014	Hanlon	03/01/85	3.95	45.0
0015	Thorton	01/03/86	3.75	45.5
0016	Millene	09/01/86	3.35	39.5
0017	May	02/01/87	0.05	30.0
0018	Malton	06/30/87	9.00	39.0
0019	Marshall	00/12/88	4.75	33.5

Figura 5-4. El contenido del fichero EMPDATA.DAT

Ahora estás preparado para decirle a STATGRAPHICS los detalles del fichero EMPDATA.DAT que se desea importar. En la pantalla de entrada de datos para Importación de ficheros, primero tienes que especificar el tipo del fichero EMPDATA.DAT. Como se dijo antes, EMPDATA.DAT es un fichero con delimitadores en blanco (véase la Figura 5-5).

- Pulsa la tecla Tab, si hace falta, para poner el cursor en el campo Input File Type (Tipo del fichero de entrada).
- Usa la barra espaciadora para conmutar este campo a Blank Delimited (delimitador en blanco).



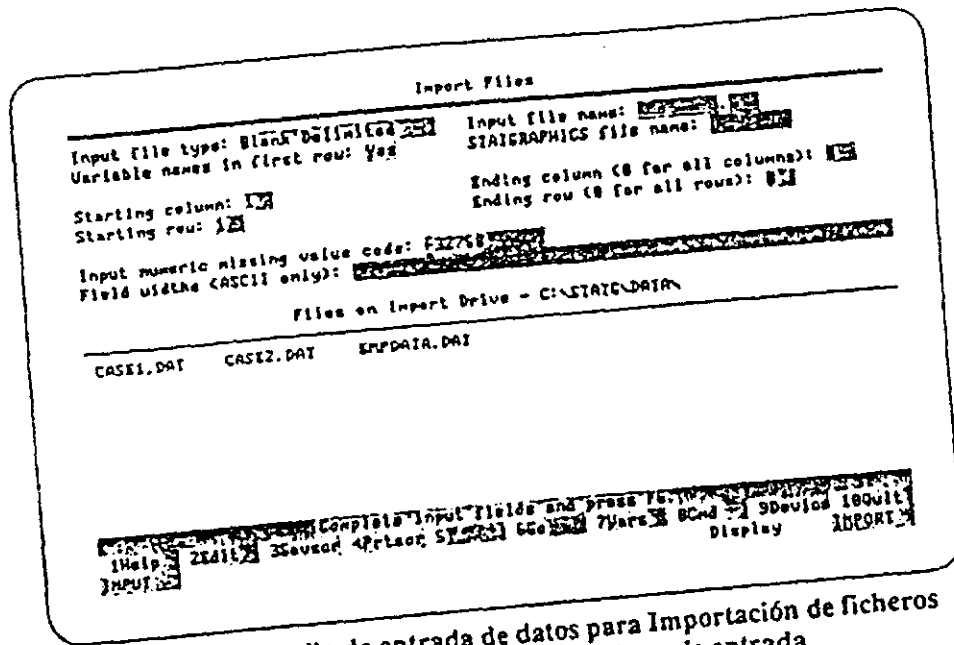


Figura 5-5. Pantalla de entrada de datos para Importación de ficheros con especificación del tipo del fichero de entrada como delimitado por blancos

Luego tienes que indicar el nombre del fichero a importar EMPDATA.DAT. (La extensión DAT en el nombre del fichero quiere decir "datos"). Se usa el campo Input File Name (Nombre del fichero de entrada) en la pantalla de entrada de datos para importación de ficheros para poner esta información (véase la Figura 5-6).

- Pulsa la tecl Tab para poner el cursor en el campo Input File Name (Nombre del fichero de entrada).
- Escribe EMPDATA en la primera parte de este campo.
- Pulsa la tecla Tab para pasar a la segunda parte del campo.
- Escribe DAT en la segunda parte del campo.

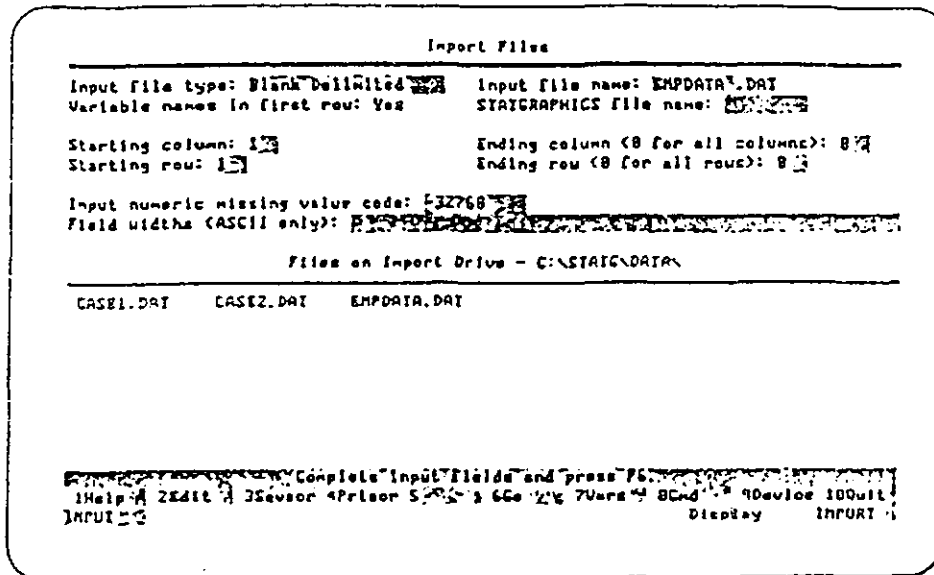


Figura 5-6. Pantalla de entrada de datos para Importación de ficheros, con EMPDATA.DAT indicado como nombre del fichero de entrada

Como viste en la Figura 5-4, la primera fila del fichero EMPDATA.DAT tiene encabezamientos de columna. Como opción por defecto, STATGRAPHICS usa los valores en la primera fila del fichero importado como nombres de las variables. (Además, como opción por defecto, STATGRAPHICS usa los valores en el resto de las filas como datos). Vamos a usar los nombres de las columnas en la primera fila del fichero EMPDATA.DAT como nombres de las variables, así que deja el campo Variable Names in First Row (Nombres de las variables en la primera fila) en Yes (Sí) (véase la Figura 5-7).

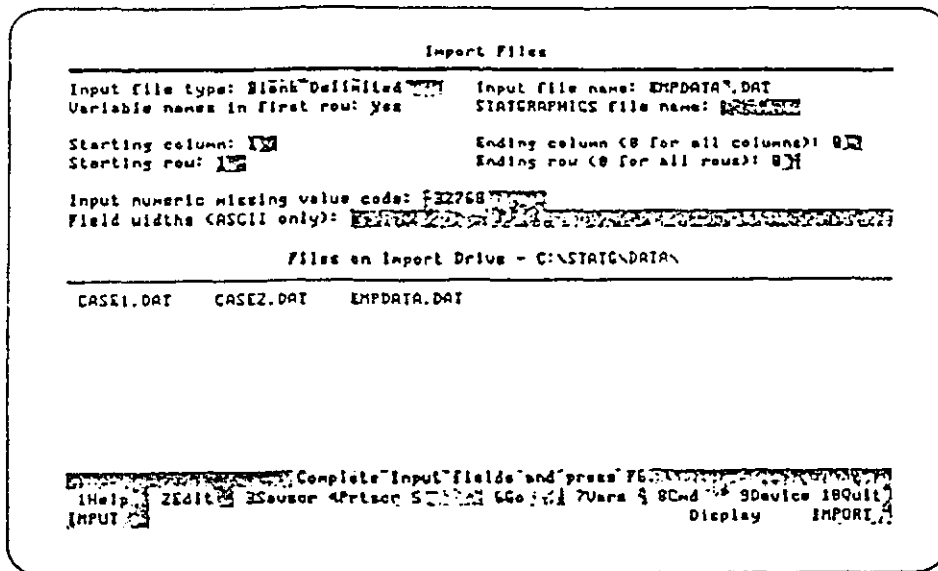


Figura 5-7. Pantalla de entrada de datos para Importación de ficheros con indicación afirmativa de Variable Names in First Row (Nombres de variables en la primera fila)

Ahora se necesita nombre para el fichero en el que se van a almacenar los datos importados. Vamos a usar como nombre de STATGRAPHICS el mismo nombre que tenía el fichero ASCII original— EMPDATA. En el campo STATGRAPHICS File Name (Nombre del fichero STATGRAPHICS) escribe EMPDATA (véase la Figura 5-8). STATGRAPHICS añade automáticamente la extensión .ASF al nombre del fichero que pongas en este campo. (ASF en realidad quiere decir APL Shared File - Fichero compartido de APL, pero puedes pensar que quiere decir A STATGRAPHICS File - un fichero de STATGRAPHICS).

Observa que tanto el campo del fichero de entrada como el campo del fichero STATGRAPHICS (de salida) indican EMPDATA. Sin embargo, el fichero de entrada tiene la extensión .DAT y el fichero de STATGRAPHICS tiene la extensión .ASF.

- Pulsa la tecla Tab para poner el cursor en el campo STATGRAPHICS File Name (Nombre del fichero STATGRAPHICS).
- Escribe el nombre del fichero EMPDATA en este campo.

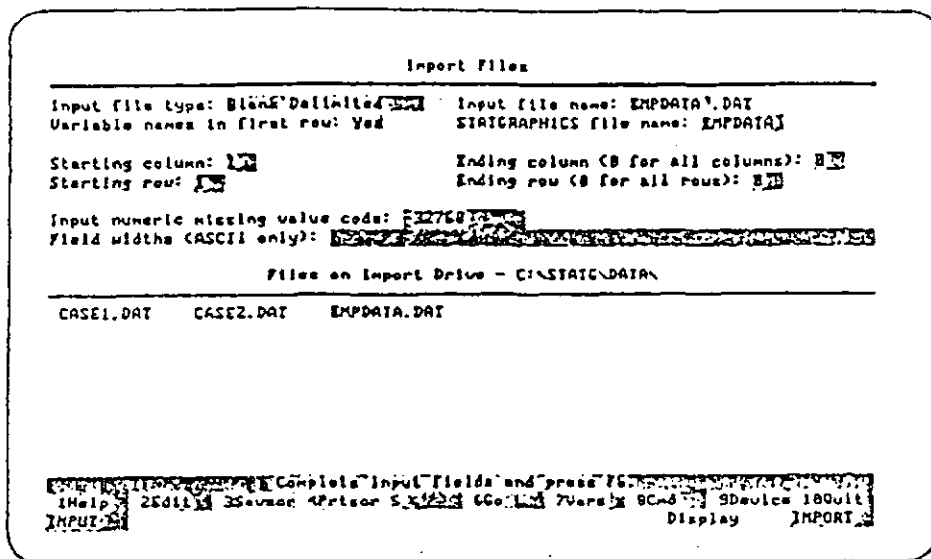


Figura 5-8. Pantalla de entrada de datos para Importación de ficheros con EMPDATA.DAT especificado como fichero de STATGRAPHICS.

Antes de importar el fichero EMPDATA.DAT, cerciórate de que las entradas en el resto de los campos sean las que aparecen en la Figura 5-9. Pasa de un campo a otro usando la tecla Tab. Para mover el cursor dentro de un campo, usa las flechas. La barra espaciadora se usa para conmutar las opciones en los campos Input File Type (Tipo del fichero de entrada) y Variable Names in First Row (Nombres de las variables en la primera fila). De ser necesario, se pueden usar las teclas normales para editar lo entrado en los campos.

Ahora estás listo para importar el fichero EMPDATA.DAT.

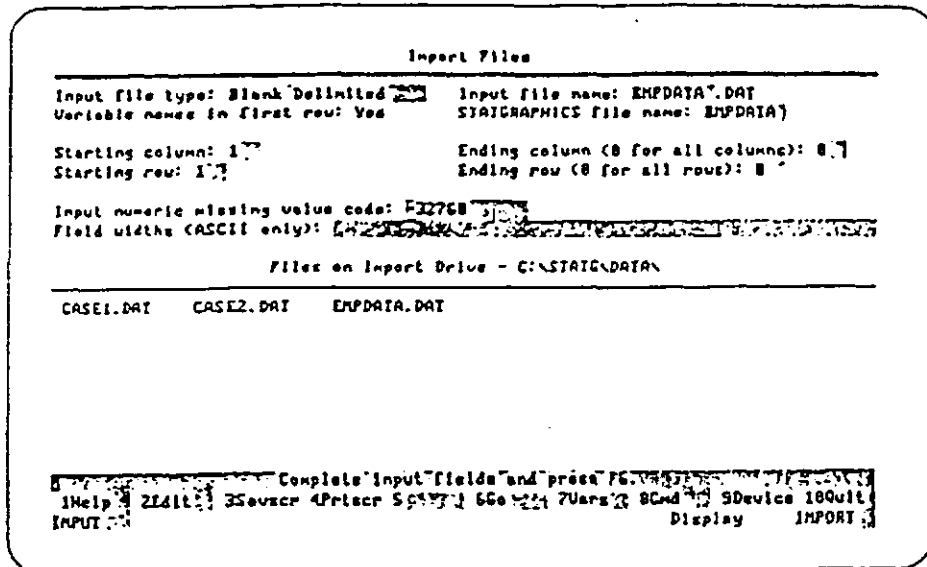


Figura 5-9. Pantalla de entrada de datos para Importación de ficheros una vez completada

- Pulsa F6 (Go).

Una vez que se pulsa F6 (Go), el sistema empieza a importar el fichero EMPDATA.DAT. Conforme va procesando el fichero, se muestra la pantalla de Información sobre la importación, que contiene detalles sobre el fichero importado, EMPDATA.DAT, y sobre el fichero STATGRAPHICS que se está creando, EMPDATA.ASF (véanse las cuatro líneas de arriba en la Figura 5-10).

Cuando se termina la importación, el sistema indica el nombre, tipo, intervalo y longitud de cada una de las variables creadas (véase el resto de las líneas de la Figura 5-10). El intervalo de una variable indica la forma de los datos: 0 es un solo número o carácter, 1 es una lista de números o de caracteres, y 2 es una tabla de filas y columnas con números o caracteres. (STATGRAPHICS importa cada columna del fichero EMPDATA.DAT como variable numérica o como variable de caracteres. Como EMPDATA.DAT es un fichero con delimitadores en blanco, STATGRAPHICS hace una variable numérica de cada columna de números, y una variable de caracteres de cada columna de números, y caracteres que no sean numéricos). El sistema indica también la fecha y la hora en que se crearon las variables.

```

Import Blank Delimited File
-----
Importing file: C:\STATC\DATA\EMPDATA.DAT
File has 28 rows and 5 columns.
Importing 28 rows and 5 columns.
Processing rows 1 to 28.
File processing complete. File Directory is:
FDIRECTORY                               11/28/98 14:49 C:\STATC\DATA\EMPDATA.DAT
ID                                         11/28/98 14:49
Name                                       11/28/98 14:49
Hire_Date                                  11/28/98 14:49
Wage_hr                                    11/28/98 14:49
Hrs_Week                                   11/28/98 14:49
-----
Press ENTER to continue.
Help 2Edit 3Saveor 4Printor 5Frc 6Col 7Vars 8Cmd 9Devios 10Quit
)APU)

```

Figura 5-10. Pantalla de Información de importación

La Figura 5-11 muestra las variables en el nuevo fichero EMPDATA.ASF. (Ya sabes cómo ver todas las variables de un fichero, así que no hace falta que lo hagas ahora. Lo puedes ver en la Figura 5-11).

Mira la anotación Typ/Wth (Tipo/amplitud) cerca del final de la pantalla. Como ves, el sistema importó las columnas del nombre y de la fecha de empleo como variables de caracteres y el resto como variables numéricas. Observa que, como se dijo antes, el sistema cambió los nombres de columna de Wage/hr y Hrs/week a los nombres de variables Wage-hr y Hrs-Week (Salario-hora y Horas-semana).

Cursor at Row: 1      Data Editor      Maximum Row: 19  
 Column: 1      File: EMPDATA      Number of Col: 5

Row	ID	Name	Hire Date	Wage_hr	Mrs_Week
1	0000	Smith	02/23/80	5.50	38.8
2	0001	Jones	03/30/80	6.25	48.0
3	0002	Spinks	07/01/80	4.75	20.0
4	0003	Fogel	09/30/80	4.55	25.8
5	0004	Whitson	01/01/81	3.95	37.5
6	0005	Wise	05/31/81	3.35	45.8
7	0006	Lindqvist	10/31/81	4.00	30.8
8	0007	Baker	04/01/82	7.10	27.5
9	0008	Duerr	12/01/82	6.75	20.8
10	0009	White	06/30/83	5.50	28.8
11	0010	Montgomery	08/15/83	5.50	24.5
12	0011	Washington	02/29/84	18.20	25.5
13	0012	Bittel	07/01/84	8.30	50.8
14	0013	Patterson	12/31/84	4.50	40.8

Length 19      19      19      19      19  
 Typ/Wth M/13      C/10      C/8      M/13      M/13

Press the F6 key to save updated data or the F5 key for other options.  
 [Help] [Edit] [Save] [Print] [50pts] [60] [70] [80] [90] [Quit] [100] [Quit]  
 INPUT      Display      FILE

Figura 5-11. Las variables en el fichero EMPDATA.ASF

Ahora que ya has terminado la lección, puedes volver al Menú principal. Desde la pantalla de Información sobre la importación se vuelve al Menú principal saliendo del procedimiento de Importación de ficheros de datos (véase la Figura 5-12).

- Pulsa la tecla Esc hasta volver al Menú principal.

STATGRAPHICS Statistical Graphics System	
<b>DATA MANAGEMENT AND SYSTEM UTILITIES</b>	<b>TIME SERIES PROCEDURES</b>
A. Data Management	L. Forecasting
B. System Environment	M. Quality Control
C. Report Writer and Graphics Replay	N. Smoothing
D. Graphics Attributes	O. Time Series Analysis
<b>PLOTTING AND DESCRIPTIVE STATISTICS</b>	<b>ADVANCED PROCEDURES</b>
E. Plotting Functions	P. Categorical Data Analysis
F. Descriptive Methods	Q. Multivariate Methods
G. Estimation and Testing	R. Nonparametric Methods
H. Distribution Functions	S. Sampling
I. Exploratory Data Analysis	T. Experimental Design
<b>HOUR AND REGRESSION ANALYSIS</b>	<b>MATHEMATICAL AND USER PROCEDURES</b>
J. Analysis of Variance	U. Mathematical Functions
K. Regression Analysis	V. Macros and User Functions

Use cursor keys to highlight desired section. Then press ENTER.  
 [Help] [Edit] [Save] [Print] [50pts] [60] [70] [80] [90] [Quit] [100] [Quit]  
 INPUT      Display

Figura 5-12. El Menú principal

## OTRAS PRACTICAS.

Para practicar más la importación de ficheros, haz lo que se sugiere a continuación.

Vuelve al procedimiento de Importación de ficheros de datos. Importa el fichero EMPDATA.DAT como fichero con formato ASCII. Indica lo siguiente: en el campo Input File Type (Tipo del fichero de entrada), ASCII; en el campo Input File Name (nombre del fichero de entrada), EMPDATA.DAT; y en el campo STATGRAPHICS File Name (nombre del fichero STATGRAPHICS), TEST.

Cuando se importa un fichero ASCII formado, hay que definir la amplitud de cada campo. Examina la Figura 5-4 para determinar la amplitud de cada columna, contando desde el primer carácter de una columna al primer carácter de la columna adyacente.

Debería resultar que la amplitud de la columna primera es 5; 11, la de la segunda ; 10, la de la tercera; 8, la de la cuarta; y 9, la de la quinta. Escribe los números 5 11 10 8 9 en el campo Field Widths (Amplitud de los campos) en la pantalla de entrada de datos para Importación de ficheros.

Pulsa **F6 (Go)** para importar el fichero. Emplea el editor de datos para examinar las variables importadas.



